PARAMETRIC INFERENCE FOR SMALL VARIANCE AND LONG TIME HORIZON MCKEAN-VLASOV DIFFUSION MODELS.

V. GENON-CATALOT[(1)], C. LARÉDO[(2)]

Abstract. Let (X_t) be solution of a one-dimensional McKean-Vlasov stochastic differential equation with classical drift term $V(\alpha, x)$, self-stabilizing term $\Phi(\beta, x)$ and small noise amplitude ε. Our aim is to study the estimation of the unknown parameters α, β from a continuous observation of $(X_t, t \in [0, T])$ under the double asymptotic framework ε tends to 0 and T tends to infinity. After centering and normalization of the process, uniform bounds for moments with respect to $t \geq 0$ and ε are derived. We then build an explicit approximate log-likelihood leading to consistent and asymptotically Gaussian estimators with original rates of convergence: the rate for the estimation of α is either ε^{-1} or $\sqrt{T}\varepsilon^{-1}$, the rate for the estimation of β is $\sqrt{T}$.

January 4, 2021

Keywords and phrases: McKean-Vlasov stochastic differential equations, Continuous observations, Small noise, Infinite time horizon, Parametric inference, Asymptotic properties of estimators.

AMS Classification. 60J60, 60J99, 62F12, 62M05

1. Introduction

We develop an approximate likelihood approach for estimating the unknown parameters of a dynamical model subject to three sources of forcing: the geometry of the state space is described by a potential term $V(\alpha, x)$, a Brownian motion with small noise allows to include small random pertubations and a self-stabilization term $\Phi(\beta, x)$. Such processes appear when describing the limit behaviour of a large population of interacting particles with an interaction function between the dynamical systems. More precisely, we study the inference for the one-dimensional process

$$dX_t = V(\alpha, X_t)dt - b(\theta, t, \varepsilon, X_t)dt + \varepsilon dW_t, \quad X_0 = x_0, \tag{1}$$

where (W_t) is a Wiener process, x_0 is deterministic known,

$$b(\theta, t, \varepsilon, x) = \int_{\mathbb{R}} \Phi(\beta, x - y) u_t^{\theta,\varepsilon}(dy), \tag{2}$$

$u_t^{\theta,\varepsilon}(dy) := u_t^{\theta,\varepsilon,x_0}(dy)$ is the distribution of $X_t := X_t^{\theta,\varepsilon,x_0}$, $V : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $\Phi : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ are deterministic Borel functions and $\theta = (\alpha, \beta) \in \Theta = \Theta_\alpha \times \Theta_\beta \subset \mathbb{R}^2$ is an unknown parameter.

(1): Université de Paris, CNRS, MAP5, UMR 8145, F-75006 Paris, France,
email: valentine.genon-catalot@parisdescartes.fr,

(2): MaiAGE, INRAE, Université Paris-Saclay, 78350 Jouy-en-Josas, France
email:catherine.laredo@inrae.fr

.

A solution of (1) is the couple $(X_t, u_t^{\theta,\varepsilon}(dy))_{t\geq 0}$ composed of the stochastic process (X_t) and the family of distributions $(u_t^{\theta,\varepsilon})$. The function $x \to b(\theta, t, \varepsilon, x)$ depends on θ, t, ε, the starting point x_0 and $u_t^{\theta,\varepsilon}$. When defined, the process (1) is a time-*inhomogeneous* Markov process known as self-stabilizing diffusion, nonlinear stochastic differential equation, or McKean-Vlasov stochastic differential equation.

These models were first described by McKean (1966) and arised in Statistical Physics for the modeling of granular media by interacting particle systems (see *e.g.* Benedetto *et al*, 1997). Due to their growing importance, many fundamental probabilistic tools for their study were developed later (see *e.g.* Gärtner (1988), Sznitman (1991) for a survey, Méléard, (1996), Benachour *et al.* (1998a, 1998b), Malrieu (2003) and many others). Herrmann *et al.* (2008) were concerned with small noise properties and large deviations results for these processes.

However, except Kasonga (1990), the statistical inference for such models remained unstudied for many years. Since 2010, the fields of application of self-stabilizing non linear differential equations progressively encompassed Statistical Physics and these equations were shown to describe collective and observable dynamics in other application fields such as Mathematical Biology (see *e.g.* Baladron *et al.* (2012)), Mogilner and Edelstein-Keshet (1999)), Epidemics Dynamics with two levels of mixing (see Ball and Sirl (2020), Forien and Pardoux (2020)), Finance (see references in Giesecke *et al.* (2020)). Several authors were concerned by statistical studies. Kasonga (1990), Gesiecke *et al.* (2020), Della Maestria and Hoffmann (2020) are interested in inference based on the direct observation of large interacting particle sytems.

Inference for stochastic differential equations (SDEs) ($\Phi(\beta, x) \equiv 0$) based on the observation of sample paths on a time interval $[0, T]$ has been widely investigated. Authors consider continuous or discrete observations, parametric or nonparametric inference under various asymptotic frameworks: small diffusion asymptotics on a fixed time interval; long time interval, especially for ergodic models; observation of n *i.i.d.* paths with large n. Among many studies, we refer first to several textbooks: Kutoyants (1984, 2004), Iacus (2010), Kessler *et al.* (2012), Höpfner (2014). Second, among the many papers on the topic, we can quote: Genon-Catalot (1990), Larédo (1990), Yoshida (1992a-b), Hoffmann (1999), Sørensen and Uchida (2003), Gobet *et al.* (2004), Dalalyan (2005), Dalalyan and Reiss (2007), Comte *et al.* (2007), Gloter and Sørensen (2009), Genon-Catalot and Larédo (2014), Guy *et al.* (2014), Comte and Genon-Catalot (2020). Moreover, these works have opened the field of inference for more complex stochastic differential equations: diffusions with jumps (see *e.g.* Masuda (2007), Schmisser (2019), Amorino and Gloter (2020)), SDEs driven by Lévy processes (see *e.g.* Masuda (2019)), diffusions with mixed effects (see *e.g.* Piccini *et al.* (2010), Delattre *et al.* (2013, 2018)), stochastic partial differential equations (see *e.g.* Cialenco (2017), Altmeyer and Reiss (2020)).

Now, the convergence as N tends to infinity of systems of N interacting particules has been investigated. One of the most important limiting processes is the class of Mc-Kean Vlasov diffusion processes (see it e.g. Sznitman (1991), Méléard (1996)). Therefore, it is a worthwhile stochastic model to study from the statistical point of view.
In Genon-Catalot and Larédo (2020), the statistical inference based on the continuous observation on a fixed time interval $[0, T]$ of (1) is investigated. Estimation of (α, β) is studied as $\varepsilon \to 0$. It appears that only α can be consistently estimated in this framework but not β. Assuming that n *i.i.d.* sample paths are observed on the fixed interval $[0, T]$, that $\varepsilon \to 0$ and n tends to infinity, both parameters are estimated but they have different rates, $\sqrt{n}\varepsilon^{-1}$ for α, $\sqrt{n}$ for β.

In the present paper, we study the parametric inference of (α, β) for the Mc-Kean Vlasov model (1) assuming that (X_t) is continuously observed on [0,T] in the new double asymptotic framework $\varepsilon \to 0, T \to \infty$. Up to our knowledge, this framework has not been studied before. As a side result, the inference for classical SDEs in this asymptotic framework is obtained.

In Section 2, we study probabilistic properties of the process (X_t). We prove that all the moments of $\varepsilon^{-1}(X_t - x_t(\alpha))$ are uniformly bounded in $t \geq 0, \varepsilon \leq 1$ (Theorem 2), that the Gaussian approximating process of (X_t) as $\varepsilon \to 0$ obtained in Genon-Catalot and Larédo (2020) holds on $\mathbb{R}^+$ and that the remainder terms of this approximation have moments uniformly bounded in $t \geq 0, \varepsilon \leq 1$ (Theorem 3). This requires some additional assumptions: we assume that there is a unique $x^*(\alpha)$ such that $V(\alpha, x^*(\alpha)) = 0$ and that this value is an attractive point for the ordinary differential equation $x_t(\alpha) = x_0 + \int_0^t V(\alpha, x_s(\alpha))ds$.
In Section 3, we define an approximate log-likelihood and study its properties together with the associated estimators as $\varepsilon \to 0$ and $T \to +\infty$ in such a way that $\varepsilon\sqrt{T} \to 0$. Two cases have to be distinguished. Either, Case **(1)**, $\frac{dx^*}{d\alpha}(\alpha) \not\equiv 0$ or Case **(2)**, $\frac{dx^*}{d\alpha}(\alpha) \equiv 0$. In the two cases, we obtain that the estimators of (α, β) are consistent. In Case **(1)**, the estimator of α is asymptotically Gaussian with the fast rate $\sqrt{T}\varepsilon^{-1}$ while in Case **(2)**, its rate is ε^{-1}. In both cases, the parameter β is estimated at rate $\sqrt{T}$. (Theorems 4-5-6). Section 4 gives some concluding remarks. Proofs are gathered in Section 5. Throughout the paper, we assume that $\varepsilon \leq 1$.

2. Probabilistic properties

2.1. Assumptions and recap of previous results.

We consider the following assumptions:

- [H0] For all α, β, the functions $x \to V(\alpha, x)$ and $x \to \Phi(\beta, x)$ are locally Lipschitz.
- [H1] Either, $\Phi(\beta, .) \equiv 0$ for all β, or for all β the function $x \to \Phi(\beta, x)$ is odd, increasing and grows at most polynomially: there exist $K(\beta) > 0$ and $r(\beta) \in \mathbb{N}$ such that $|\Phi(\beta, x) - \Phi(\beta, y)| \leq |x - y|(K(\beta) + |x|^{r(\beta)} + |y|^{r(\beta)}),\ x, y \in \mathbb{R}$.
- [H2-k] The functions $x \to V(\alpha, x)$ and $x \to \Phi(\beta, x)$ have continuous partial derivatives up to order k and these derivatives have polynomial growth: for all α, β, and all $i, i \leq k$, there exist constants $k(\alpha) > 0, k(\beta) > 0$ and integers $\gamma(\alpha) \geq 0, \gamma(\beta) \geq 0$, such that
$$\forall x \in \mathbb{R},\ |\frac{\partial^i V}{\partial x^i}(\alpha, x)| \leq k(\alpha)(1 + |x|^{\gamma(\alpha)}), \quad |\frac{\partial^i \Phi}{\partial x^i}(\beta, x)| \leq k(\beta)(1 + |x|^{\gamma(\beta)}).$$
- [H3] For all α, the function $x \to V(\alpha, x)$ is continuously differentiable and there exists $K_V(\alpha) > 0$ such that $\forall x \in \mathbb{R}, \quad \frac{\partial V}{\partial x}(\alpha, x) \leq -K_V(\alpha)$.
- [H4] There exists $x^*(\alpha)$ such that $V(\alpha, x^*(\alpha)) = 0$.

Note that the case $\Phi(\beta, x) \equiv 0$ corresponds to a classical stochastic differential equation which under [H3]-[H4] admits a unique invariant distribution.

Let us recall some results of Herrmann *et al.* (2008) where Equation (1) is studied in the more general case of X_0 a random variable, independent of (W_t) with distribution μ. Under [H0]-[H1] and [H3], if $\mathbb{E}X_0^{8q^2} < +\infty$ where $q = [(r(\beta)/2) + 1]$, then, for all θ, there exists a drift term $b(\theta, t, \varepsilon, x) = b^\mu(\theta, t, \varepsilon, x)$ such that (1) admits a unique strong solution $(X_t = X_t^{\theta,\varepsilon,\mu})$ satisfying $b(\theta, t, \varepsilon, x) = \int_{\mathbb{R}} \Phi(\beta, x - y)u_t^{\theta,\varepsilon,\mu}(dy)$ and X is the unique strong solution of (1). Moreover, for all $n \in \{1, \ldots, 4q^2\}$, whenever $\mathbb{E}X_0^{2n} < +\infty$, $\sup_{t\geq 0} \mathbb{E}X_t^{2n} < +\infty$. Since we assume here that $X_0 = x_0$ is deterministic, this yields that, for all $n \in \mathbb{N}$, $\sup_{t\geq 0} \mathbb{E}X_t^{2n} < +\infty$. Under [H3], $x^*(\alpha)$ in [H4] is the unique value such that $V(\alpha, x^*(\alpha)) = 0$.
Under [H0]-[H1] and [H3], the process (X_t) admits a unique invariant distribution for all fixed

$\varepsilon > 0$ (see *e.g.* Cattiaux *et al.* (2008)).

In a previous paper, we have studied the process (X_t) on a fixed time interval $[0,T]$. Let us recall the results that we need in the sequel. First, properties of continuity and differentiability of $b(\theta,t,\varepsilon,x)$ defined in (2) with respect to ε and x at $(\theta,t,0,x)$ can be derived from the assumptions.

Lemma 1. *(Genon-Catalot and Larédo (2020)) Assume [H0]-[H1], [H2-2], [H3] and that $X_0 = x_0$ is deterministic. Then*
(i) For all $\theta, t \geq 0$, $(\varepsilon, x) \to b(\theta,t,\varepsilon,x)$ is continuously differentiable on $[0,+\infty) \times \mathbb{R}$.
(ii) $\lim_{\varepsilon \to 0} b(\theta,t,\varepsilon,x) = \Phi(\beta, x - x_t(\alpha))$.
(iii) At $\varepsilon = 0$, $\frac{\partial b}{\partial \varepsilon}(\theta,t,0,x) = 0$ and $\frac{\partial b}{\partial x}(\theta,t,0,x) = \frac{\partial \phi}{\partial x}(\beta, x - x_t(\alpha))$.

Property (ii) is also proved in Hermann *et al.* (2008).
Next, the asymptotic properties of (X_t) on a fixed time interval $[0,T]$ as $\varepsilon \to 0$ have been studied. Consider the ordinary differential equation associated to $\varepsilon = 0$.

$$dx_t(\alpha) = V(\alpha, x_t(\alpha))dt, \; x_0(\alpha) = x_0. \tag{3}$$

As ε tends to 0, (X_t) converges uniformly in probability on $[0,T]$ to $x_t(\alpha)$. Moreover, setting

$$a(\theta,t) = \frac{\partial V}{\partial x}(\alpha, x_t(\alpha)) - \frac{\partial \Phi}{\partial x}(\beta, 0), \tag{4}$$

define $(g_t(\theta))$ the Ornstein-Uhlenbeck process

$$dg_t(\theta) = a(\theta,t)g_t(\theta)dt + dW_t, \quad g_0(\theta) = 0. \tag{5}$$

Note that $\frac{\partial \Phi}{\partial x}(\beta,0) \geq 0$ so that, under [H3], $a(\theta,t) \leq -(K_V(\alpha) + \frac{\partial \Phi}{\partial x}(\beta,0)) < 0$. Then, the following expansion of (X_t) with respect to ε holds.

Theorem 1. *(Genon-Catalot and Larédo (2020)) Assume [H0], [H1] and [H2-3], then*

$$X_t = x_t(\alpha) + \varepsilon g_t(\theta) + \varepsilon^2 R_t^\varepsilon(\theta), \tag{6}$$

where the remainder term $R_t^\varepsilon(\theta)$ has moments uniformly bounded on $[0,T]$.

Equation (5) can be solved

$$g_t(\theta) = \int_0^t \exp\Big(\int_s^t a(\theta,u)du\Big)dW_s = \int_0^t e^{A(\theta,t)-A(\theta,s)}dW_s, \text{ where} \tag{7}$$

$$A(\theta,t) = \int_0^t a(\theta,u)du. \tag{8}$$

In order to illustrate the results, we considered the following explicit example.

Example 1. Consider the model where $V(\alpha,x) = -\alpha\, x, \Phi(\beta,x) = \beta\, x$ with $\alpha > 0, \beta \geq 0$. We have $b(\theta,t,\varepsilon,x) = \beta(x - \mathbb{E}_\theta(X_t))$, and equation (1) writes:

$$dX_t = -\alpha X_t dt - \beta(X_t - \mathbb{E}_\theta(X_t))dt + \varepsilon dW_t, \quad X_0 = x_0.$$

We easily check that $\mathbb{E}_\theta(X_t) = x_0 e^{-\alpha t}$ and (1) can be solved explicitely:

$$X_t = x_0 e^{-\alpha t} + \varepsilon e^{-(\alpha+\beta)t}\int_0^t e^{(\alpha+\beta)s}dW_s. \tag{9}$$

The remainder term $R_t^\varepsilon(\theta)$ is here equal to 0.

2.2. **Statement of probabilistic results.** Under the assumptions of Section 2.1, we can extend the previous results and prove uniform bounds on $\mathbb{R}^+$.

Theorem 2. *Let (X_t) denote the solution of (1) and $x_t(\alpha)$ the solution of (3).*
(i) *Assume [H0]-[H1], [H3]. Then, for all $n \geq 1$, there exists a constant $\delta(\alpha, n)$ such that*

$$\forall \varepsilon \in (0,1], \quad \forall t \geq 0, \quad \mathbb{E}_\theta \left(\frac{X_t - x_t(\alpha)}{\varepsilon} \right)^{2n} \leq \delta(\alpha, n).$$

(ii) *If moreover [H2-2] and [H4] hold, there exists a constant $\delta(\alpha) > 0$ such that,*

$$\forall \varepsilon \in (0,1] \ \ \forall t \geq 0, \ \ \varepsilon^{-2} |\mathbb{E}_\theta(X_t - x_t(\alpha))| \leq \delta(\alpha).$$

In the special case where $V(\alpha, x) = -\alpha x$, $\mathbb{E}_\theta(X_t) = x_t(\alpha)$.

Remark 1. *We can easily check that, under [H0], [H3]-[H4],*

$$(x_t(\alpha) - x^*(\alpha))^2 \leq (x_0 - x^*(\alpha))^2 \exp(-2K_V(\alpha)t). \tag{10}$$

Therefore $x_t(\alpha)$ converges as $t \to +\infty$ to $x^*(\alpha)$ with exponential rate.

Remark 2. *It follows immediately from Theorem 2 and Remark 1 that, under [H0]-[H1], [H3]-[H4], $X_t \to x^*(\alpha)$ in probability as $t \to +\infty$ and $\varepsilon \to 0$. Therefore the Dirac measure $\delta_{x^*(\alpha)}$ appears as the limit of the distribution of (X_t) as $t \to +\infty$.*

Remark 3. *The bounds $\delta(\alpha, n)$ and $\delta(\alpha)$ depend on θ only through α. From the proofs, we have that $\delta(\alpha, n) = (nK_V^{-1}(\alpha))^n$ and that $\delta(\alpha)$ is a function of $B(\alpha), k(\alpha), \gamma(\alpha), K_V^{-1}(\alpha)$, where $B(\alpha) = \sup_{t \geq 0} |x_t(\alpha)| < +\infty$. These bounds increase in each of its variables (see (33) and (36)). Thus, if these quantities are upper bounded by constants independent of α, the bounds of Theorem 2 are uniform in α.*

We also have that the remainder term $R_t^\varepsilon(\theta)$ defined in (6) has moments uniformly bounded on $\mathbb{R}^+$.

Theorem 3. *Under [H0]-[H1], [H2-3], [H3]-[H4], the expansion $X_t = x_t(\alpha) + \varepsilon g_t(\theta) + \varepsilon^2 R_t^\varepsilon(\theta)$ holds on $\mathbb{R}^+$ and $R_t^\varepsilon(\theta)$ satisfies*

$$\sup_{t \geq 0, \varepsilon \in (0,1]} \mathbb{E}_\theta |R_t^\varepsilon(\theta)| = O(1) \ \text{ and for all } \ p \geq 1, \ \sup_{t \geq 0, \varepsilon \in (0,1]} \mathbb{E}_\theta (R_t^\varepsilon(\theta) - \mathbb{E}_\theta R_t^\varepsilon(\theta))^{2p} = O(1).$$

Remark 4. *Note that $\mathbb{E}_\theta g_t^2(\theta)$ is uniformly bounded on $\mathbb{R}^+$. Indeed, using the explicit expression of $g_t(\theta)$ given in (7) and the property that, under [H3], for $s \leq t$, $A(\theta, t) - A(\theta, s) \leq -K_V(\alpha)(t - s)$, we get $\mathbb{E}_\theta g_t^2(\theta) = \int_0^t \exp[2(A(\theta, t) - A(\theta, s))] ds \ \leq (2K_V(\alpha))^{-1}$.*

Define

$$D(\theta, t, \varepsilon, x) = b(\theta, t, \varepsilon, x) - \Phi(\beta, x - x_t(\alpha)). \tag{11}$$

The following corollary dealing with $D(\theta, t, \varepsilon, X_t)$ is a crucial tool for the statistical study. As for $R_t^\varepsilon(\theta)$, uniform bounds hold for $D(\theta, t, \varepsilon, X_t)$.

Corollary 1. *Assume [H0]-[H1], [H2-3], [H3]-[H4]. Then $D(\theta, t, \varepsilon, X_t)$ defined in (11) satisfies,*

$$\sup_{t \in \mathbb{R}^+, \, \varepsilon \in (0,1]} \varepsilon^{-2} \, |\mathbb{E}_\theta D(\theta, t, \varepsilon, X_t)| = O(1), \tag{12}$$

$$\forall p \geq 1, \quad \sup_{t \in \mathbb{R}^+, \, \varepsilon \in (0,1]} \varepsilon^{-6p} \mathbb{E}_\theta (D(\theta, t, \varepsilon, X_t) - \mathbb{E}_\theta D(\theta, t, \varepsilon, X_t))^{2p} = O(1). \tag{13}$$

Remark 5. *The constants $O(1)$ in Theorem 3 and Corollary 1 are independent of θ if the constants $k(\alpha), k(\beta), \gamma(\alpha), \gamma(\beta), K_V^{-1}(\alpha), B(\alpha)$ are upper bounded independently of θ (see Remark 3).*

3. Estimation when both ε tends to 0 and T tends to infinity

As it is usual in statistics, we consider the canonical space associated with the observation of $(X_t)_{t\in[0,T]}$, $(\Omega, \mathcal{F}, (\mathcal{F}_t, t \in [0,T]), \mathbb{P}_\theta)$, where $\Omega = C([0,T])$ is the space of continuous real-valued functions defined on $[0,T]$ endowed with the Borel σ-field associated with the uniform convergence on $[0,T]$, $(X_t, t \in [0,T])$ is the canonical process $(X_t(\omega) = \omega(t))$, $(\mathcal{F}_t, t \in [0,T])$ is the canonical filtration and $\mathbb{P}_\theta$ is the distribution of (1) on $C([0,T])$.
In this section, we study the estimation of (α, β) from a continuous observation $(X_t, t \in [0,T])$ and, in addition to [H0]-[H1], [H3]-[H4], we assume

- [S0] The parameter set is $\Theta = \Theta_\alpha \times \Theta_\beta$ where $\Theta_\alpha, \Theta_\beta$ are bounded closed intervals. The true value of the parameter is $\theta_0 = (\alpha_0, \beta_0)$ and belongs to $\mathring{\Theta}$.
- [S1] The function $(\alpha, x) \to V(\alpha, x)$ (resp. $(\beta, x) \to \Phi(\beta, x)$) is defined and continuous on $U_\alpha \times \mathbb{R}$ (resp. $U_\beta \times \mathbb{R}$), and all the derivatives $(\alpha, x) \to \dfrac{\partial^{i+j} V}{\partial x^i \partial \alpha^j}(\alpha, x)$, $(\beta, x) \to \dfrac{\partial^{i+j}\Phi}{\partial x^i \partial \beta^j}(\beta, x)$ exist, are continuous on $U_\alpha \times \mathbb{R}$ (resp. $U_\beta \times \mathbb{R}$), where U_α, U_β are open intervals containing respectively $\Theta_\alpha, \Theta_\beta$, and have polynomial growth with respect to x: there exist a constant $K > 0$ and a nonnegative integer k such that
$$\forall (\alpha,\beta) \in \Theta, \forall x \in \mathbb{R}, \forall i, j \geq 0, \ |\frac{\partial^{i+j} V}{\partial x^i \partial \alpha^j}(\alpha, x)| + |\frac{\partial^{i+j}\Phi}{\partial x^i \partial \beta^j}(\beta, x)| \leq K(1 + |x|^k).$$
- [S2] There exists $K_V > 0$ such that : $\forall \alpha \in \Theta_\alpha, K_V(\alpha) \geq K_V > 0$. (see [H3])

Assumption [S0] is standard in parametric inference and used only for consistency. Assuming the existence of derivatives of any order is not necessary but it simplifies the exposure.
The uniformity of the constants K, k, K_V in [S1]-[S2] is only required for the consistency part. As $\Theta_\alpha, \Theta_\beta$ are supposed to be compact, this is not a strong assumption.

By the relation $V(\alpha, x^*(\alpha)) = 0$, the function $\alpha \to x^*(\alpha)$ is continuous so, as Θ_α is compact, $\sup_{\alpha\in\Theta_\alpha} |x^*(\alpha) = A < +\infty$. Therefore, under [S2], $\sup_{\alpha\in\Theta_\alpha} \sup_{t\geq 0} |x_t(\alpha)| = B < +\infty$. In view of Remarks 3 and 5, under [S1]-[S2], all the bounds of Theorems 2, 3 and Corollary 1 are not only uniform in t, ε but also in θ.

3.1. **Approximate likelihood.** The Girsanov formula holds for nonlinear self-stabilizing diffusions and the log-likelihood associated with the observation of $(X_t, t \in [0,T])$ is
$$\ell_{\varepsilon,T}(\theta) = \frac{1}{\varepsilon^2}\int_0^T (V(\alpha, X_s) - b(\theta, s, \varepsilon, X_s))\, dX_s - \frac{1}{2\varepsilon^2}\int_0^T (V(\alpha, X_s) - b(\theta, s, \varepsilon, X_s))^2\, ds.$$
It contains the term $b(\theta, s, \varepsilon, X_s)$ which is involved for the estimation of θ. However, for small ε, $b(\theta, t, \varepsilon, X_t) = \int \Phi(\beta, x - y) u_t^{\theta,\varepsilon}(dy)$ is close to $\Phi(\beta, x - x_t(\alpha))$ (see Lemma 1). Therefore, as in Genon-Catalot and Larédo (2020), we consider an approximate log-likelihood where we replace $b(\theta, s, \varepsilon, x)$ by $\Phi(\beta, x - x_t(\alpha))$ and set
$$\Lambda_{\varepsilon,T}(\theta) = \frac{1}{\varepsilon^2}\int_0^T H(\theta, s, X_s) dX_s - \frac{1}{2\varepsilon^2}\int_0^T H^2(\theta, s, X_s) ds, \quad \text{with} \tag{14}$$
$$H(\theta, s, x) = V(\alpha, x) - \Phi(\beta, x - x_s(\alpha)). \tag{15}$$
This approximate log-likelihood is easier to study. We have previously obtained that asymptotic efficiency is kept for α with this approximate log-likelihood and that β cannot be estimated on a fixed time interval $[0,T]$ (see Genon-Catalot and Larédo (2020)). Therefore, to estimate both parameters, we have to combine two asymptotic frameworks. In Genon-Catalot and Larédo

(2020), we considered n *i.i.d.* paths of process (1) with $\varepsilon \to 0$ and $n \to +\infty$. Here, we investigate, for the observation of one path, the combination of $\varepsilon \to 0$ and $T \to +\infty$.

3.2. **Preliminary results.** Let us set

$$\ell(\alpha) = -\frac{\partial V}{\partial x}(\alpha, x^*(\alpha)) \geq K_V(\alpha) \geq K_V > 0, \quad \ell(\alpha,\beta) = \ell(\alpha) + \frac{\partial \Phi}{\partial x}(\beta, 0) \geq \ell(\alpha). \tag{16}$$

Proposition 1. *Assume [H0], [H3],[H4]. The triplet* $(x_t(\alpha), \frac{\partial x_t}{\partial \alpha}(\alpha,t), \frac{\partial^2 x_t}{\partial \alpha^2}(\alpha,t))$ *converges to* $(x^*(\alpha), \frac{dx^*}{d\alpha}(\alpha), \frac{d^2x^*}{d\alpha^2}(\alpha))$ *exponentially fast with rate* $\exp(-\ell(\alpha)t)$ *as* t *tends to infinity.*

We also need to specify the asymptotic behaviour of functionals of the time inhomogeneous process $(g_t(\theta))$ defined by(5) in Theorem 1 or by (7).

Proposition 2. *Assume [H0]-[H1], [H3]-[H4]. Then,* $(g_t(\theta))$ *satisfies as* $T \to \infty$,

(i) $\frac{1}{T}\int_0^T [g_t(\theta)]^2 dt \to_{\mathbb{L}^2(\mathbb{P}_\theta)} [2\ell(\alpha,\beta)]^{-1}$,
(ii) $\frac{1}{T}\int_0^T g_t(\theta) dt \to_{\mathbb{P}_\theta} 0$,
(iii) *If the function* $h : \mathbb{R}^+ \to \mathbb{R}^+$ *satisfies* $\lim_{t\to+\infty} h(t) = 0$, *then* $\frac{1}{\sqrt{T}}\int_0^T g_t(\theta)h(t)dt \to_{\mathbb{P}_\theta} 0$.

3.3. **Rates of convergence.** We may now study the joint estimation of (α, β).
The asymptotic distribution of (X_t) as $t \to \infty$ and $\varepsilon \to 0$ is the Dirac measure $\delta_{x^*(\alpha)}$ (see Remark 2). As detailed below, the estimation of α varies according to the property that $x^*(\alpha)$ depends on α or not. Indeed, for all α, $V(\alpha, x^*(\alpha)) \equiv 0$, thus

$$\frac{d}{d\alpha}(V(\alpha, x^*(\alpha)) = \frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) + \frac{\partial V}{\partial x}(\alpha, x^*(\alpha))\frac{dx^*}{d\alpha}(\alpha) \equiv 0. \tag{17}$$

Now, by [H3], $\frac{\partial V}{\partial x}(\alpha, x^*(\alpha)) = -\ell(\alpha) \not\equiv 0$. Therefore two cases are to be distinguished:

(1) $\frac{dx^*}{d\alpha}(\alpha) \not\equiv 0 \Leftrightarrow \frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) \not\equiv 0$.
(2) $\frac{dx^*}{d\alpha}(\alpha) \equiv 0 \Leftrightarrow \frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) \equiv 0$: $x^*(\alpha) = x^*$ does not depend on α.

Let us remark that Example 1 presented in Section 2.1 belongs to Case **(2)**.

According to these two cases, we set

$$D^{(1)}_{\varepsilon,T} = \begin{pmatrix} \frac{\varepsilon}{\sqrt{T}} & 0 \\ 0 & \frac{1}{\sqrt{T}} \end{pmatrix}, \quad D^{(2)}_{\varepsilon,T} = \begin{pmatrix} \varepsilon & 0 \\ 0 & \frac{1}{\sqrt{T}} \end{pmatrix}, \quad \mathcal{J}_{\varepsilon,T}(\theta) = -\begin{pmatrix} \frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \alpha^2}(\theta) & \frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \beta \partial \alpha}(\theta) \\ \frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \beta \partial \alpha}(\theta) & \frac{\partial^2 \Lambda_{\varepsilon,T}}{\partial \beta^2}(\theta) \end{pmatrix}. \tag{18}$$

Theorem 4. Case (1) *(*$\frac{dx^*}{d\alpha}(\alpha) \not\equiv 0$*). Assume [H0]-[H1], [H3]-[H4], [S1] and* $\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta, 0) \neq 0$. *Then, if* $\varepsilon \to 0, T \to +\infty$ *in such a way that* $\varepsilon\sqrt{T} \to 0$, *the following holds: under* $\mathbb{P}_\theta$,

$$D^{(1)}_{\varepsilon,T} \begin{pmatrix} \frac{\partial \Lambda_{\varepsilon,T}}{\partial \alpha}(\theta) \\ \\ \frac{\partial \Lambda_{\varepsilon,T}}{\partial \beta}(\theta) \end{pmatrix} = \begin{pmatrix} \frac{\varepsilon}{\sqrt{T}} \frac{\partial \Lambda_{\varepsilon,T}}{\partial \alpha}(\theta) \\ \\ \frac{1}{\sqrt{T}} \frac{\partial \Lambda_{\varepsilon,T}}{\partial \beta}(\theta) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(\mathbf{0}, \mathcal{J}^{(1)}(\theta)), \tag{19}$$

where

$$\mathcal{J}^{(1)}(\theta) = \begin{pmatrix} \ell^2(\alpha,\beta)\left(\frac{\frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha))}{\ell(\alpha)}\right)^2 & 0 \\ \\ 0 & \frac{\left(\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta,0)\right)^2}{2\ell(\alpha,\beta)} \end{pmatrix} \tag{20}$$

and $\ell(\alpha), \ell(\alpha,\beta)$ *are defined in* (16). *Moreover, the matrix* $-D^{(1)}_{\varepsilon,T}\mathcal{J}_{\varepsilon,T}(\theta)D^{(1)}_{\varepsilon,T} = \mathcal{J}^{(1)}(\theta) + o_P(1)$.

Theorem 5. Case (2) *($\forall \alpha, x^*(\alpha) = x^*$). Assume [H0]-[H1], [H3]-[H4], [S1], $\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0) \neq 0$ and $x_0 \neq x^*$. If $\varepsilon \to 0, T \to +\infty$ in such a way that $\varepsilon\sqrt{T} \to 0$, then under $\mathbb{P}_\theta$,*

$$D^{(2)}_{\varepsilon,T}\begin{pmatrix} \frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta) \\ \\ \frac{\partial\Lambda_{\varepsilon,T}}{\partial\beta}(\theta)\end{pmatrix} = \begin{pmatrix} \varepsilon\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta) \\ \\ \frac{1}{\sqrt{T}}\frac{\partial\Lambda_{\varepsilon,T}}{\partial\beta}(\theta)\end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(\mathbf{0}, \mathcal{J}^{(2)}(\theta)), \tag{21}$$

where

$$\mathcal{J}^{(2)}(\theta) = \begin{pmatrix} \int_0^{+\infty}[\frac{\partial V}{\partial\alpha}(\alpha, x_s(\alpha)) + \frac{\partial\Phi}{\partial x}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha,s))]^2 ds & 0 \\ 0 & \frac{\left(\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\right)^2}{2\ell(\alpha,\beta)}\end{pmatrix} \tag{22}$$

and $\ell(\alpha,\beta)$ is defined in (16). *Moreover, the matrix $-D^{(2)}_{\varepsilon,T}\mathcal{J}_{\varepsilon,T}(\theta)D^{(2)}_{\varepsilon,T} = \mathcal{J}^{(2)}(\theta) + o_P(1)$.*

In Theorem 5, the additional condition $x_0 \neq x^*$ appears as a minimal assumption. Indeed, since x^* does not depend on α, $x_0 = x^*$ implies that, for all α and all $s \geq 0$, $x_s(\alpha) = x^*$ and, using (17), the term $\mathcal{J}^{(2)}(\theta)_{11} = 0$. In Case **(2)**, the integrand in $\mathcal{J}^{(2)}(\theta)_{11}$ tends to 0 as s tends to ∞. This convergence is exponential (see Proposition 1), so that $\mathcal{J}^{(2)}(\theta)_{11}$ is finite.
We stress that Theorems 4 and 5 show that the estimation of α and β have different rates of convergence. While in both cases, β is estimated at rate $\sqrt{T}$, according to the assumptions α is estimated at rate $\sqrt{T}\varepsilon^{-1}$ or ε^{-1}.
We can check that these rates hold also for α when $\Phi(\beta,.) \equiv 0$,(*i.e.* for classical stochastic differential equations), the condition $\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0) \neq 0$ being required only for β. This yields the corollary stated below.

Corollary 2. *Assume that $\Phi(\beta,.) \equiv 0$ (classical stochastic differential equation) and [H0], [H3],[H4] and [S1]. The contrast $\Lambda_{\varepsilon,T}(\theta)$ is equal to the exact log-likelihood $\ell_{\varepsilon,T}(\alpha)$ (it depends only on α). Then, if $\varepsilon \to 0, T \to +\infty$ in such a way that $\varepsilon\sqrt{T} \to 0$, the following holds:*
If $\frac{dx^*}{d\alpha}(\alpha) \not\equiv 0$, under $\mathbb{P}_\alpha$,

$$\frac{\varepsilon}{\sqrt{T}}\ell'_T(\alpha) \to_{\mathcal{D}} \mathcal{N}(0, \left[\frac{\partial V}{\partial\alpha}(\alpha, x^*(\alpha))\right]^2), \quad \frac{\varepsilon^2}{T}\ell''_T(\alpha) \to -\left[\frac{\partial V}{\partial\alpha}(\alpha, x^*(\alpha))\right]^2.$$

If $\frac{dx^*}{d\alpha}(\alpha) \equiv 0$ and $x_0 \neq x^*$, under $\mathbb{P}_\alpha$,

$$\varepsilon\ell'_T(\alpha) \to_{\mathcal{D}} \mathcal{N}(0, \int_0^{+\infty}\left[\frac{\partial V}{\partial\alpha}(\alpha, x_s(\alpha))\right]^2 ds), \quad \varepsilon^2\ell''_T(\alpha) \to -\int_0^{+\infty}\left[\frac{\partial V}{\partial\alpha}(\alpha, x_s(\alpha))\right]^2 ds.$$

Up to our knowledge, these statistical results are also new for classical stochastic differential equations. Indeed, for ergodic diffusion processes with fixed diffusion term ε, the rate of convergence for α is $\sqrt{T}$ as T tends to infinity, while on a fixed time interval $[0,T]$, as ε tends to 0, the rate of estimation for α is ε^{-1}. With the double asymptotics $\varepsilon \to 0$ and $T \to +\infty$, it is unexpected to obtain a rate of convergence for α which is either $\varepsilon^{-1}\sqrt{T}$ or ε^{-1}. This distinction depends on the fact that the fixed point $x^*(\alpha)$ of the ODE depends on α or not.

3.4. **Asymptotic properties of estimators.**

Consider the approximate likelihood $\Lambda_{\varepsilon,T}$ defined in (14),(15) and let $(\hat{\alpha}_{\varepsilon,T}, \hat{\beta}_{\varepsilon,T})$ denote the maximum pseudo-likelihood estimator defined as any solution of

$$(\hat{\alpha}_{\varepsilon,T}, \hat{\beta}_{\varepsilon,T}) = \arg\max_{(\alpha,\beta)\in\Theta_\alpha\times\Theta_\beta} \Lambda_{\varepsilon,T}(\alpha,\beta). \tag{23}$$

Define the three functions using (16),

$$\Lambda_1^{(1)}(\alpha,\alpha_0,\beta) = -\frac{1}{2}\left(V(\alpha,x^*(\alpha_0)) - \Phi(\beta,x^*(\alpha_0)-x^*(\alpha))\right)^2, \tag{24}$$

$$\Lambda_1^{(2)}(\alpha,\alpha_0,\beta) = -\frac{1}{2}\int_0^{+\infty}\left[V(\alpha,x_s(\alpha_0)) - V(\alpha_0,x_s(\alpha_0)) - \Phi(\beta,x_s(\alpha_0)-x_s(\alpha))\right]^2 ds, \tag{25}$$

$$\Lambda_2(\alpha_0,\beta,\beta_0) = -\frac{1}{2}\left(\frac{\partial\Phi}{\partial x}(\beta,0) - \frac{\partial\Phi}{\partial x}(\beta_0,0)\right)^2 \frac{1}{2\ell(\alpha_0,\beta_0)}. \tag{26}$$

Lemma 2. *Assume [H0], [H1], [H3], [H4], [S0], [S1]. Then, as $\varepsilon\to 0$ and $T\to+\infty$ in such a way that $\varepsilon\sqrt{T}\to 0$, the following holds in probability under $\mathbb{P}_{\theta_0}$:*
(i) Case (1) *($\frac{dx^*}{d\alpha}(\alpha)\not\equiv 0$). Uniformly with respect to $(\alpha,\beta)\in\Theta_\alpha\times\Theta_\beta$,*
$\frac{\varepsilon^2}{T}(\Lambda_{\varepsilon,T}(\alpha,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta))\to\Lambda_1^{(1)}(\alpha,\alpha_0,\beta)$.
(ii) Case (2) *($\frac{dx^*}{d\alpha}(\alpha)\equiv 0$). Uniformly with respect to $(\alpha,\beta)\in\Theta_\alpha\times\Theta_\beta$,*
$\varepsilon^2(\Lambda_{\varepsilon,T}(\alpha,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta))\to\Lambda_1^{(2)}(\alpha,\alpha_0,\beta)$.
(iii) Both cases. *Uniformly with respect to $\beta\in\Theta_\beta$*
$\frac{1}{T}(\Lambda_{\varepsilon,T}(\alpha_0,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta_0))\to\Lambda_2(\alpha_0,\beta,\beta_0)$.

Let us determine the identifiability assumptions associated with Lemma 2.
Case **(1)**: Assume that $\forall\beta,\ \Lambda_1^{(1)}(\alpha,\alpha_0,\beta)=0$. This implies

$$\forall\beta,\ \ V(\alpha,x^*(\alpha_0)) = \Phi(\beta,x^*(\alpha_0)-x^*(\alpha)).$$

As $\Phi(\beta,.)$ is an increasing function, this yields that

$$V(\alpha,x^*(\alpha_0)) = 0 \quad\text{and}\quad \Phi(\beta,x^*(\alpha_0)-x^*(\alpha)) = 0.$$

Since $\Phi(\beta,x)=0$ implies $x=0$, the last equality implies $x^*(\alpha)=x^*(\alpha_0)$.
Consider now the case of standard SDE : $\Phi(\beta,.)\equiv 0$. The condition $V(\alpha,x^*(\alpha_0))=0$ implies $x^*(\alpha)=x^*(\alpha_0)$ by the uniqueness of the fixed point.
For β, the identifiability assumption is straightforward since $\ell(\alpha_0,\beta_0)>0$.
Therefore, we deduce the identifiability assumptions for Case **(1)**:

- [S3] $x^*(\alpha)=x^*(\alpha_0)\Rightarrow\alpha=\alpha_0$.
- [S4] $\frac{\partial\Phi}{\partial x}(\beta,0)=\frac{\partial\Phi}{\partial x}(\beta_0,0)\Rightarrow\beta=\beta_0$.

Consider now Case **(2)** where for all α, $x^*(\alpha)=x^*$.
If $x_0=x^*$, $x_s(\alpha_0)=x_s(\alpha)=x^*$ for all $s\geq 0$. Thus, $\Lambda_1^{(2)}(\alpha,\alpha_0,\beta)=0$.
Assume now that $x_0\neq x^*$. The term under the integral in (25) converges to 0 exponentially fast (see Proposition 1). Hence, $\Lambda_1^{(2)}(\alpha,\alpha_0,\beta)$ is well defined and finite. This leads to the following identifiability assumption of α in Case (2):

- [S5] $x_0\neq x^*$; $\{s\to V(\alpha,x_s(\alpha_0))-V(\alpha_0,x_s(\alpha_0))\equiv 0$ and $s\to x_s(\alpha)-x_s(\alpha_0)\equiv 0\}\Rightarrow\{\alpha=\alpha_0\}$.

Nothing changes for β.
If $\Phi(\beta,.)\equiv 0$, then [S5] has to be changed into:
[*S5b*]: $x_0\neq x^*$; $\{s\to V(\alpha,x_s(\alpha_0))-V(\alpha_0,x_s(\alpha_0))\equiv 0\Rightarrow\{\alpha=\alpha_0\}$.

The inference for (α,β) is a two-rate statistical problem. According to Gloter and Sørensen (2009, Section 4.4.1), the proof of the consistency of $(\hat\alpha_{\varepsilon,T},\hat\beta_{\varepsilon,T})$ relies on three steps: (1) Prove that $\hat\alpha_{\varepsilon,T}$ is consistent; (2) Prove that $\varepsilon^{-1}\sqrt{T}(\hat\alpha_{\varepsilon,T}-\alpha_0)$ in Case **(1)**, $\varepsilon^{-1}(\hat\alpha_{\varepsilon,T}-\alpha_0)$ in Case

(2)) is $\mathbb{P}_{\theta_0}$- tight; (3) Prove that $\hat{\beta}_{\varepsilon,T}$ is consistent. The proof of the asymptotic normality is then obtained standardly based on Theorems 4 and 5.

Theorem 6. *Assume [H0], [H1], [H3], [H4], [S0], [S1] and that $\varepsilon \to 0$ and $T \to +\infty$ in such a way that $\varepsilon\sqrt{T} \to 0$.*
Case (1) *($\frac{dx^*}{d\alpha}(\alpha) \not\equiv 0$). Assume moreover that [S3]-[S4] hold. Then $(\hat{\alpha}_{\varepsilon,T}, \hat{\beta}_{\varepsilon,T})$ is consistent and, under $\mathbb{P}_{\theta_0}$,*

$$\begin{pmatrix} \frac{\sqrt{T}}{\varepsilon}(\hat{\alpha}_{\varepsilon,T} - \alpha_0) \\ \sqrt{T}(\hat{\beta}_{\varepsilon,T} - \beta_0) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(0, [\mathcal{J}^{(1)}(\theta_0)]^{-1}), \quad \text{where } \mathcal{J}^{1)}(\theta) \text{ is defined in (20).}$$

Case (2) *($\forall \alpha, x^*(\alpha) = x^*$). Assume moreover that [S4]-[S5] hold. Then $(\hat{\alpha}_{\varepsilon,T}, \hat{\beta}_{\varepsilon,T})$ is consistent and, under $\mathbb{P}_{\theta_0}$,*

$$\begin{pmatrix} \frac{1}{\varepsilon}(\hat{\alpha}_{\varepsilon,T} - \alpha_0) \\ \sqrt{T}(\hat{\beta}_{\varepsilon,T} - \beta_0) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(0, [\mathcal{J}^{(2)}(\theta_0)]^{-1}), \quad \text{where } \mathcal{J}^{(2)}(\theta) \text{ is defined in (22).}$$

Let us consider simple examples that illustrate these results.

Example 1 (continued): Let $V(\alpha, x) = -\alpha x$, $\Phi(\beta, x) = \beta x$, $\alpha > 0, \beta > 0$.
As $x^*(\alpha) = x^* = 0$, we are in Case **(2)**. The contrast is equal to the exact log-likelihood, and as $\varepsilon \to 0$, $T \to +\infty$ with $\varepsilon\sqrt{T} \to 0$, applying Theorem 5 yields that, under $\mathbb{P}_\theta$,

$$\begin{pmatrix} \varepsilon \frac{\partial \ell_{\varepsilon,T}}{\partial \alpha}(\theta) \\ \frac{1}{\sqrt{T}} \frac{\partial \ell_{\varepsilon,T}}{\partial \beta}(\theta) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(\mathbf{0}, \mathcal{J}^{(2)}(\theta)) \text{ with } \quad \mathcal{J}^{(2)}(\theta) = \begin{pmatrix} x_0^2 \int_0^\infty (1+s\beta)^2 e^{-2\alpha s} ds & 0 \\ 0 & \frac{1}{2(\alpha+\beta)} \end{pmatrix}.$$

The functions $\Lambda_1^{(2)}(\alpha, \alpha_0, \beta)$ and $\Lambda_2(\alpha_0, \beta, \beta_0)$ are explicit.

$$\Lambda_1^{(2)}(\alpha, \alpha_0, \beta) = -x_0^2 \, \frac{(\alpha+\beta)^2 + \alpha_0\alpha}{4\alpha_0\alpha(\alpha_0+\alpha)}(\alpha - \alpha_0)^2, \;\; \Lambda_2(\alpha_0, \beta, \beta_0) = -\frac{(\beta - \beta_0)^2}{4(\alpha_0 + \beta_0)}.$$

Hence, the two identifiability assumptions [S4], [S5] are satisfied and

$$\begin{pmatrix} \frac{1}{\varepsilon}(\hat{\alpha}_{\varepsilon,T} - \alpha_0) \\ \sqrt{T}(\hat{\beta}_{\varepsilon,T} - \beta_0) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(0, \mathcal{J}^{(2)}(\theta_0)^{-1}).$$

Example 2 : Consider the slightly different case $V(\alpha, x) = -\alpha x + 1$, $\Phi(\beta, x) = \beta x$, $\alpha > 0, \beta > 0$.
We have $x^*(\alpha) = \alpha^{-1}$. We are in Case **(1)** : $\frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) = -\alpha^{-1} \neq 0$. Applying Theorem 4 to $\Lambda_{\varepsilon,T}$, which is the exact log-likelihood

$$\begin{pmatrix} \frac{\varepsilon}{\sqrt{T}} \frac{\partial \ell_{\varepsilon,T}}{\partial \alpha}(\theta) \\ \frac{1}{\sqrt{T}} \frac{\partial \ell_{\varepsilon,T}}{\partial \beta}(\theta) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(\mathbf{0}, \mathcal{J}^{(1)}(\theta)) \text{ with } \quad \mathcal{J}^{(1)}(\theta) = \begin{pmatrix} \frac{(\alpha+\beta)^2}{\alpha^4} & 0 \\ 0 & \frac{1}{2(\alpha+\beta)} \end{pmatrix}.$$

We have:

$$\Lambda_1^{(1)}(\alpha, \alpha_0, \beta) = -\frac{(\alpha+\beta)^2}{2\alpha_0^2\alpha^2}(\alpha - \alpha_0)^2, \;\; \Lambda_2(\alpha_0, \beta, \beta_0) = -\frac{(\beta - \beta_0)^2}{4(\alpha_0 + \beta_0)}.$$

Hence, the two identifiability assumptions [S4], [S5] are satisfied and

$$\begin{pmatrix} \frac{\sqrt{T}}{\varepsilon}(\hat{\alpha}_{\varepsilon,T} - \alpha_0) \\ \sqrt{T}(\hat{\beta}_{\varepsilon,T} - \beta_0) \end{pmatrix} \to_{\mathcal{L}} \mathcal{N}_2(0, \mathcal{J}^{(1)}(\theta_0)^{-1}).$$

Example 3 : Let $V(\alpha,x)=-\alpha x$ and arbitrary Φ. We are in Case **(2)**, $x^*(\alpha)=x^*=0$, $\ell(\alpha)=\alpha$ and

$$\Lambda_1^{(2)}(\alpha,\alpha_0,\beta)=-\frac{1}{2}\int_0^\infty [x_0(\alpha_0-\alpha)e^{-\alpha_0 s}-\phi(\beta,x_0(e^{-\alpha_0 s}-e^{-\alpha s}))]^2 ds$$

$$\Lambda_2(\alpha_0,\beta,\beta_0)=-\frac{1}{2}(\frac{\partial\Phi}{\partial x}(\beta,0)-\frac{\partial\Phi}{\partial x}(\beta_0,0))^2\times\frac{1}{2(\alpha_0+\frac{\partial\Phi}{\partial x}(\beta_0,0))}. \tag{27}$$

By [S4]-[S5], $\Lambda_1^{(2)}(\alpha,\alpha_0,\beta)=0$ iff $x_0=x^*$ or $\alpha=\alpha_0$ and $\Lambda_2(\alpha_0,\beta,\beta_0)=0$ iff $\beta=\beta_0$. Theorem 6 implies the consistency and the asymptotic normality of Case **(2)**.

Example 4: For $\Phi(\beta,x)=\beta(x^3+x)$, $\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)=1$. Therefore, β can be estimated. But if $\Phi(\beta,x)=\beta x^3$, then $\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)=0$, and by this method, we cannot estimate β.

4. Concluding remarks

In this paper, we consider the one-dimensional McKean-Vlasov process (X_t) given by (1) with small noise ε, under assumptions ensuring existence and uniqueness of solutions. We are interested in the statistical estimation of the unknown parameters α,β present in the classical drift term $V(\alpha,x)$ and in the self-stabilizing term $\Phi(\beta,x)$. In a previous paper (Genon-Catalot and Larédo (2020)), we have shown that, on the basis of one trajectory continuously observed on a time interval $[0,T]$, while it is possible to estimate consistently α as ε tends to 0, it is not possible to estimate β if T is kept fixed. This is why in this paper, we consider the double asymptotic framework $\varepsilon\to 0$ and $T\to+\infty$. This requires some additional assumptions on the model ([H3]-[H4]). In particular, we assume that there is a unique $x^*(\alpha)$ such that $V(\alpha,x^*(\alpha))=0$ and this value is an attractive point for the ordinary differential equation $x_t(\alpha)=x_0+\int_0^t V(\alpha,x_s(\alpha))ds$. We stress that this double asymptotic framework has never been studied even for classical stochastic differential equations (corresponding to $\Phi(\beta,.)\equiv 0$).

In a first part, we study probabilistic properties of the process (X_t). We prove that all the moments of $\varepsilon^{-1}(X_t-x_t(\alpha))$ are uniformly bounded in $t\geq 0,\varepsilon\leq 1$, that the Gaussian approximating process of (X_t) as $\varepsilon\to 0$ obtained in Genon-Catalot and Larédo (2020) holds on $\mathbb{R}^+$ and that the remainder terms of this approximation have moments uniformly bounded in $t\geq 0,\varepsilon\leq 1$.

In a second part, we define a contrast (approximate log-likelihood) and prove the consistency and asymptotic normality of the corresponding maximum contrast estimators as $\varepsilon\to 0$ and $T\to+\infty$ in such a way that $\varepsilon\sqrt{T}\to 0$. For the estimation of α, two cases have to be distinguished. Either Case **(1)**, $\frac{dx^*}{d\alpha}(\alpha)\not\equiv 0$ or Case **(2)**, $\frac{dx^*}{d\alpha}(\alpha)\equiv 0$. In Case **(1)**, the estimator of α is asymptotically Gaussian with the fast rate $\sqrt{T}\varepsilon^{-1}$ while in Case **(2)**, its rate is ε^{-1}. In both cases, the parameter β is estimated at rate $\sqrt{T}$. This confirms the fact that a double asymptotic is needed for estimating both α and β on the basis of one trajectory.

We did not study here the asymptotic efficiency of our estimators since this can be proved as in our previous paper (Genon-Catalot and Larédo, 2020, Section 6) by means of an asymptotic equivalence of experiments property.

Extensions of this work could be to consider multidimensional Mc-Kean Vlasov models of the more general form (see *e.g.* Méléard (1996), Sznitmann (1991):

$$dX_t=b(\theta,t,X_t,u_t^{\theta,c})dt+\varepsilon\sigma(c,t,X_t,u_t^{\theta,c})dW_t.$$

where θ,c is are unknown parameters, $u_t^{\theta,c}$ is the distribution of X_t.

Another direction would be to study, for discrete observations of McKean-Vlasov diffusions,

the estimation of both parameters in the drift and in the diffusion coefficient as in Gloter and Sorensen (2009) and Guy *et al.* (2014).

5. Appendix

5.1. Proofs of Section 2.

Proof of Theorem 2. Let

$$\zeta_t = (X_t - x_t(\alpha)/\varepsilon \quad \text{with distribution } \nu_t^\varepsilon(dz). \tag{28}$$

We have

$$d\zeta_t = \varepsilon^{-1}(V(\alpha,X_t) - V(\alpha,x_t(\alpha)))dt - \frac{1}{\varepsilon}\left(\int \Phi(\beta,X_t-y)u_t^{\theta,\varepsilon}(dy)\right)dt + dW_t, \text{ so that}$$

$$\zeta_t^2 = 2\int_0^t \zeta_s dW_s + 2\int_0^t \zeta_s\, \varepsilon^{-1}[V(\alpha,X_s) - V(\alpha,x_s(\alpha)]ds - 2\varepsilon^{-1}\int_0^t \zeta_s\left(\int \Phi(\beta,X_s-y)u_s^{\theta,\varepsilon}(dy)\right)ds + t.$$

This implies, setting $m_2^\varepsilon(t) := \mathbb{E}_\theta\zeta_t^2$,

$$m_2^\varepsilon(t) = 2\int_0^t \mathbb{E}_\theta[\zeta_s\frac{1}{\varepsilon}(V(\alpha,X_s) - V(\alpha,x_s(\alpha)))]ds - \frac{2}{\varepsilon}\int_0^t \mathbb{E}_\theta[\zeta_s\int \Phi(\beta,X_s-y)u_s^{\theta,\varepsilon}(dy)]ds + t. \tag{29}$$

Using (28) and $\Phi(\beta,X_s-y) = \Phi(\beta,X_s - x_s - (y-x_s))$, we get

$$\begin{aligned} 2\mathbb{E}_\theta\left(\zeta_s\varepsilon^{-1}\int \Phi(\beta,X_s-y)u_s^{\theta,\varepsilon}(dy)\right) &= \frac{2}{\varepsilon}\int z\int \Phi(\beta,\varepsilon(z-z'))\varepsilon\nu_s^\varepsilon(dz)\nu_s^\varepsilon(dz') \\ &= \int (z-z')\int \Phi(\beta,\varepsilon(z-z'))\nu_s^\varepsilon(dz)\nu_s^\varepsilon(dz') \geq 0. \end{aligned}$$

Therefore, since $\Phi(\beta,X_s-y) = \Phi(\beta,\varepsilon(\zeta_s - \frac{y-x_s}{\varepsilon})$, we get

$$\begin{aligned} 2\mathbb{E}_\theta\left(\zeta_s\varepsilon^{-1}\int \Phi(\beta,X_s-y)u_s(dy)\right) &= 2\varepsilon^{-1}\int z\int \Phi(\beta,\varepsilon(z-z'))\varepsilon\nu_s^\varepsilon(dz')\ \nu_s^\varepsilon(dz) \\ &= \int (z-z')\int \Phi(\beta,\varepsilon(z-z'))\nu_s^\varepsilon(dz)\nu_s^\varepsilon(dz') \geq 0. \end{aligned}$$

Differentiating (29) and using [H3], we get $(m_2^\varepsilon)'(t) \leq -2K_V(\alpha)m_2^\varepsilon(t) + 1$.
Now, we can use the following property which holds for $f(.)$ a $C^1(\mathbb{R}^+,\mathbb{R})$ function: If there exists $\ell > 0$ such that

$$\{t\geq 0, f(t) > \ell\} \subset \{t\geq 0, f'(t) < 0\} \quad \text{then} \quad \sup_{t\geq 0} f(t) \leq f(0)\vee\ell. \tag{30}$$

Thus, choosing $\ell = \frac{1}{2K_V(\alpha)}$ yields, since $m_2^\varepsilon(0) = 0$,

$$\sup_{t\geq 0} m_2^\varepsilon(t) \leq \frac{1}{2K_V(\alpha)}. \tag{31}$$

Let us now study $m_{2n}^\varepsilon(t) := \mathbb{E}_\theta\zeta_t^{2n}$. We have

$$\zeta_t^{2n} = 2n\int_0^t \zeta_s^{2n-1}d\zeta_s + n(2n-1)\int_0^t \zeta_s^{2n-2}ds. \tag{32}$$

Analogously, for $n\geq 1$, using that $\Phi(\beta,.)$ is odd,

$$\begin{aligned} 2n\mathbb{E}_\theta\left(\zeta_s^{2n-1}\frac{1}{\varepsilon}\int \Phi(\beta,X_s-y)u_s^{\theta,\varepsilon}(dy)\right) &= 2n\varepsilon^{-1}\int z^{2n-1}\int \Phi(\beta,\varepsilon(z-z'))\varepsilon\nu_s^\varepsilon(dz)\nu_s^\varepsilon(dz') \\ &= n\int (z^{2n-1} - z'^{2n-1})\int \Phi(\beta,\varepsilon(z-z'))\nu_s^\varepsilon(dz)\nu_s^\varepsilon(dz') \geq 0. \end{aligned}$$

The first term of $d\zeta_s$ in (32) satisfies under [H3],

$$\mathbb{E}_\theta[\zeta_s^{2n-1}\frac{1}{\varepsilon}(V(\alpha,X_s)-V(\alpha,x_s(\alpha)))]=\mathbb{E}_\theta[\zeta_s^{2n-2}\zeta_s\frac{1}{\varepsilon}(V(\alpha,X_s)-V(x_s(\alpha)))]\leq -K_V(\alpha)\mathbb{E}_\theta\zeta_s^{2n}.$$

Therefore, applying the Hölder inequality to $f(x)=x^{1-1/n}$, we get

$$(m_{2n}^\varepsilon(t))'\leq -2nK_V(\alpha)\ m_{2n}^\varepsilon(t)+n(2n-1)m_{2n-2}^\varepsilon(t)\leq -2nK_V(\alpha)\ m_{2n}^\varepsilon(t)+n(2n-1)(m_{2n}^\varepsilon(t))^{1-1/n}.$$

Choosing $\delta(\alpha,n)=(\frac{n}{K_V(\alpha)})^n$, we have that, for $x\geq\delta(\alpha,n)$, $-2nK_V(\alpha)x+n(2n-1)x^{1-1/n}<0$. Thus, as $m_{2n}^\varepsilon(0)=0$, applying (30) yields

$$\sup_{t\geq 0} m_{2n}^\varepsilon(t)\leq\delta(\alpha,n), \tag{33}$$

where $\delta(\alpha,n)$ does not depend on ε and β.

It remains to study $\mathbb{E}_\theta X_t-x_t(\alpha)$. We have,

$$\mathbb{E}_\theta(X_t-x_t(\alpha))=\int_0^t\mathbb{E}_\theta\left(V(\alpha,X_s)-V(\alpha,x_s(\alpha))\right)ds-\int_0^t\mathbb{E}_\theta b(\theta,s,\varepsilon,X_s)ds.$$

Let $(\overline{X}_s)$ be an independent copy of (X_s). Then,

$$\mathbb{E}_\theta b(\theta,s,\varepsilon,X_s)=\mathbb{E}_\theta\int\Phi(\beta,X_s-y)u_s^{\theta,\varepsilon}(dy)=\mathbb{E}_\theta(\Phi(\beta,X_s-\overline{X}_s))=0, \tag{34}$$

since $\Phi(\beta,.)$ is odd and since the distribution of $X_s-\overline{X}_s$ is symmetric.
Now, a Taylor expansion at $x_s(\alpha)$ yields

$$\mathbb{E}_\theta(V(\alpha,X_s)-V(\alpha,x_s(\alpha))=\mathbb{E}_\theta(X_s-x_s(\alpha))\frac{\partial V}{\partial x}(\alpha,x_s(\alpha))+R_s,\ \text{with}$$

$$R_s=\int_0^1(1-u)\mathbb{E}_\theta\left((X_s-x_s(\alpha))^2\frac{\partial^2V}{\partial x^2}(\alpha,x_s(\alpha)+u(X_s-x_s(\alpha)))\right)du.$$

Therefore,

$$f(\theta,t):=\mathbb{E}_\theta X_t-x_t(\alpha)=\int_0^t\left(\mathbb{E}_\theta X_s-x_s(\alpha)\right)\frac{\partial V}{\partial x}(\alpha,x_s(\alpha))ds+\int_0^t R_s\,ds.$$

Differentiating with respect to t, we get that $\frac{\partial f}{\partial t}(\theta,t)=\frac{\partial V}{\partial x}(\alpha,x_t(\alpha))f(\theta,t)+R_t,\ f(\theta,0)=0$. Consequently,

$$f(\theta,t)=\int_0^t R_s\exp\left(\int_s^t\frac{\partial V}{\partial x}(\alpha,x_u(\alpha))du\right)ds. \tag{35}$$

Using [H2-2], $|R_s|\leq k(\alpha)\mathbb{E}_\theta\left((X_s-x_s(\alpha))^2(1+|x_s(\alpha)|^{\gamma(\alpha)}+|X_s-x_s(\alpha)|^{\gamma(\alpha)})\right)$.
Under [H3], [H4], $x_t(\alpha)$ is uniformly bounded on $\mathbb{R}^+$ by $B(\alpha)$ (see Remark 1). Using the first part, $\mathbb{E}_\theta(X_t-x_t(\alpha))^2\leq\frac{\varepsilon^2}{2K_V(\alpha)}$. By the Hölder inequality,

$$\mathbb{E}_\theta|X_t-x_t(\alpha)|^{2+\gamma(\alpha)}\leq\varepsilon^{2+\gamma(\alpha)}(m_{2+2\gamma(\alpha)}^\varepsilon(t))^{1-\frac{\gamma(\alpha)}{2+2\gamma(\alpha)}}.$$

Therefore, for $\varepsilon\leq 1$, $|R_s|\leq\varepsilon^2C(\alpha)$ where $C(\alpha)=\frac{k(\alpha)}{2K_V(\alpha)}(1+B(\alpha)^{\gamma(\alpha)}+2(1+\gamma(\alpha))(\frac{1+\gamma(\alpha)}{K_V(\alpha)})^{\frac{\gamma(\alpha)}{1+\gamma(\alpha)}})$ is independent of t,ε. Hence,

$$|f(\theta,t)|=|\mathbb{E}_\theta X_t-x_t(\alpha)|\leq\varepsilon^2C(\alpha)\int_0^t\exp\left(-K_V(\alpha)(t-s)\right)ds\leq\frac{C(\alpha)}{K_V(\alpha)}\,\varepsilon^2. \tag{36}$$

□

If $V(\alpha,x)=-\alpha x$, then $\mathbb{E}_\theta X_t = x_0 - \alpha\int_0^t \mathbb{E}_\theta X_s ds$, thus $\mathbb{E}_\theta X_t = x_0\exp(-\alpha t) = x_t(\alpha)$.

Proof of Theorem 3 .

By (6), we have $R_t^\varepsilon(\theta) = \varepsilon^{-2}(X_t - x_t(\alpha) - \varepsilon g_t(\theta))$. Therefore, using (1), (3), (4) and (5),

$$\begin{aligned} dR_t^\varepsilon(\theta) &= \frac{1}{\varepsilon^2}\left(V(\alpha,X_t) - V(\alpha,x_t(\alpha)) - b(\theta,t,\varepsilon,X_t) - \varepsilon a(\theta,t)g_t(\theta)\right)dt \\ &= \frac{1}{\varepsilon^2}\left[(X_t - x_t(\alpha) - \varepsilon a(\theta,t)g_t(\theta))\right]dt + \nu(\theta,t,\varepsilon,X_t)dt \\ &= a(\theta,t)R_t^\varepsilon(\theta)dt + \nu(\theta,t,\varepsilon,X_t)dt, \quad R_0^\varepsilon(\theta)=0, \text{ where} \end{aligned}$$

$$\nu(\theta,t,\varepsilon,X_t) = \frac{1}{\varepsilon^2}\left((V(\alpha,X_t) - V(\alpha,x_t(\alpha)) - b(\theta,t,\varepsilon,X_t)) - (X_t - x_t(\alpha))a(\theta,t)\right) = T_1(t)+T_2(t),$$

with $T_1(t) = \varepsilon^{-2}\left(V(\alpha,X_t) - V(\alpha,x_t(\alpha)) - (X_t - x_t(\alpha))\frac{\partial V}{\partial x}(\alpha,x_t(\alpha))\right)$,

$$T_2(t) = -\varepsilon^{-2}\left(\int \Phi(\beta,X_t-y)u_t^{\theta,\varepsilon}(dy) - \frac{\partial\Phi}{\partial x}(\beta,0)(X_t - x_t(\alpha))\right). \tag{37}$$

The equation satisfied by $R_t^\varepsilon(\theta)$ can be solved and we get, using (4) and (7),

$$R_t^\varepsilon(\theta) = \int_0^t \nu(\theta,s,\varepsilon,X_s)\exp\left(\int_s^t a(\theta,u)du\right)ds. \tag{38}$$

Let us first study $T_1(t)$. A Taylor expansion at point $x_t(\alpha)$ yields, using Assumption [H2-2],

$$\begin{aligned} T_1(t) &= \varepsilon^{-2}(X_t - x_t(\alpha))^2\int_0^1 (1-u)\frac{\partial^2 V}{\partial x^2}(\alpha, x_t(\alpha) + u(X_t - x_t(\alpha)))du, \qquad (39)\\ |T_1(t)| &\le k(\alpha)\varepsilon^{-2}(X_t - x_t(\alpha))^2(1 + |x_t(\alpha)|^{\gamma(\alpha)} + |X_t - x_t(\alpha)|^{\gamma(\alpha)}). \end{aligned}$$

Therefore, since $x_t(\alpha)$ is uniformly bounded, applying Theorem 2 yields that, for all $p\ge 1$,

$$\mathbb{E}_\theta|T_1(t)|^{2p} = O(1) \text{ uniformly on } t\ge 0, \varepsilon>0.$$

For $T_2(t)$, we have $-\varepsilon^2 T_2(t) = \int\left(\Phi(\beta,X_t-y) - \frac{\partial\Phi}{\partial x}(\beta,0)(X_t - x_t(\alpha))\right)u_t^{\theta,\varepsilon}(dy)$.
A Taylor expansion at point 0 yields, noting that $\frac{\partial^2\Phi}{\partial x^2}(\beta,0)=0$,

$$\Phi(\beta,X_t-y) - \frac{\partial\Phi}{\partial x}(\beta,0)(X_t - x_t(\alpha)) = \frac{\partial\Phi}{\partial x}(\beta,0)(x_t(\alpha) - y) + \rho_1(X_t,y),$$

where

$$\rho_1(X_t,y) = \frac{1}{2}(X_t-y)^3\int_0^1 (1-u)^2\frac{\partial^3\Phi}{\partial x^3}(\beta,u(X_t-y))du. \tag{40}$$

Therefore, $T_2(t) = -\varepsilon^{-2}\frac{\partial\Phi}{\partial x}(\beta,0))(x_t(\alpha) - \mathbb{E}_\theta X_t) - \varepsilon^{-2}\int\rho_1(X_t,y)u_t^{\theta,\varepsilon}(dy) = T_{21}(t) + T_{22}(t)$.
Let us study first $\mathbb{E}_\theta T_2(t) = T_{21}(t) + \mathbb{E}_\theta T_{22}(t)$. For the second term, we can write, for $\overline{X}_t$ an independent copy of X_t,

$$\mathbb{E}_\theta\int\rho_1(X_t,y)u_t^{\theta,\varepsilon}(dy) = \frac{1}{2}\mathbb{E}_\theta\left((X_t - \overline{X}_t)^3\int_0^1(1-u)^2\frac{\partial^3\Phi}{\partial x^3}(\beta,u(X_t-\overline{X}_t))du\right).$$

Under [H2-3], $x\to x^3\frac{\partial^3\Phi}{\partial x^3}(\beta,ux)$ is well defined and odd so that

$$\mathbb{E}_\theta\int\rho_1(X_t,y)u_t^{\theta,\varepsilon}(dy) = \mathbb{E}_\theta(\rho_1(X_t,\overline{X}_t)) = 0. \tag{41}$$

Therefore, $\mathbb{E}_\theta T_{22}(t) = 0$. For $T_{21}(t)$ which is deterministic, applying Theorem 2 (ii) yields

$$|T_{21}(t)| \leq \delta(\alpha)\frac{\partial \Phi}{\partial x}(\beta, 0) = O(1) \text{ uniformly on } t \geq 0, \varepsilon > 0.$$

Therefore $|\mathbb{E}_\theta T_2(t)| = |T_{21}(t)|$ is also uniformly bounded for $t \geq 0, \varepsilon > 0$.
Consider now $T_2(t) - \mathbb{E}_\theta T_2(t)$. Using (41), it is equal to $T_{22}(t) = -\varepsilon^{-2}\int \rho_1(X_t, y))u_t^{\theta,\varepsilon}(dy)$. Hence, if $(\bar{X}_t)$ is an independent copy of (X_t),

$$\mathbb{E}_\theta(T_2(t) - \mathbb{E}_\theta T_2(t))^{2p} = \varepsilon^{-4p}\mathbb{E}_\theta(\rho_1(X_t, \bar{X}_t)^{2p}).$$

Now, by [H2-3] and (40)

$$\begin{aligned}
\mathbb{E}_\theta(\rho_1(X_t, \bar{X}_t)^{2p}) &= 2^{-2p}\mathbb{E}_\theta\left((X_t - \overline{X}_t)^{6p}(\int_0^1 (1-u)^2 \frac{\partial^3 \Phi}{\partial x^3}(\beta, u(X_t - \overline{X}_t))du)^{2p}\right) \\
&\leq 2^{-2p}\mathbb{E}_\theta\left(|X_t - \overline{X}_t|^{6p}(k(\beta)(1 + |X_t - \overline{X}_t|^{\gamma(\beta)}))^{2p}\right) \\
&\leq k^{2p}(\beta)2^{-1}\mathbb{E}_\theta\left(|X_t - \overline{X}_t|^{6p}(1 + |X_t - \overline{X}_t|^{2p\gamma(\beta)}))\right).
\end{aligned}$$

By splitting $X_t - \overline{X}_t$ into $X_t - x_t(\alpha) + x_t(\alpha)) - \overline{X}_t$ we get that

$$\begin{aligned}
\mathbb{E}_\theta(\rho_1(X_t, \bar{X}_t)^{2p}) &\leq k^{2p}(\beta)\left(\mathbb{E}_\theta(2^{6p-1}(X_t - x_t(\alpha))^{6p}) + \mathbb{E}_\theta(2^{6p+2p\gamma(\beta)-1}(X_t - x_t(\alpha))^{6p+2p\gamma(\beta)})\right) \\
&\leq C_p(\alpha, \beta)\varepsilon^{6p},
\end{aligned}$$

where $C_p(\alpha, \beta)$ depends on p, $k(\beta)$ and $K_V^{-1}(\alpha)$. Applying Theorem 2 yields that, uniformly on $t > 0$, $\varepsilon > 0$,

$$\mathbb{E}_\theta\left(T_2(t) - \mathbb{E}_\theta T_2(t)\right)^{2p} \leq \varepsilon^{2p} C_p'(\alpha, \beta).$$

Joining these inequalities there exist constants $\delta(\alpha, \beta), \delta_p(\alpha, \beta)$ such that for all $t \geq 0, \varepsilon > 0$,

$$\mathbb{E}_\theta|\nu(\theta, t, \varepsilon, X_t)| \leq \delta(\alpha, \beta); \quad \mathbb{E}_\theta\left(\nu(\theta, t, \varepsilon, X_t) - \mathbb{E}_\theta\nu(\theta, t, \varepsilon, X_t)\right)^{2p} \leq \delta_p(\alpha, \beta).$$

Now, using (16), (8) and [H3], $\int_s^t a(\theta, u)du = A(\theta, t) - A(\theta, s) \leq -K(\theta)(t - s)$ with

$$K(\theta) = K(\alpha, \beta) = K_V(\alpha) + \frac{\partial \Phi}{\partial x}(\beta, 0) > 0. \tag{42}$$

Therefore (38) yields that

$$\mathbb{E}_\theta|R_t^\varepsilon(\theta)| \leq \int_0^t \mathbb{E}_\theta|\nu(\theta, s, \varepsilon, X_s)|e^{-K(\theta)(t-s)}ds \leq \frac{\delta(\alpha, \beta)}{K(\theta)}.$$

Consider now $\mathbb{E}_\theta(R_t^\varepsilon(\theta) - \mathbb{E}_\theta R_t^\varepsilon(\theta))^{2p}$. Equation (38) yields

$$(R_t^\varepsilon(\theta) - \mathbb{E}_\theta R_t^\varepsilon(\theta))^{2p} \leq \int_0^t \left(\nu(\theta, t, \varepsilon, X_t) - \mathbb{E}_\theta\nu(\theta, t, \varepsilon, X_t)\right)^{2p} e^{p(A(\theta,t) - A(\theta,s))}ds \left(\int_0^t e^{\frac{p}{2p-1}(A(\theta,t) - A(\theta,s))}ds\right)^{2p-1}$$

This yields, using the inequality for $A(\theta, s)$, that

$$\mathbb{E}_\theta(R_t^\varepsilon(\theta) - \mathbb{E}_\theta R_t^\varepsilon(\theta))^{2p} \leq (\frac{2p-1}{2pK(\theta)})^{2p-1}\int_0^t \mathbb{E}_\theta(\nu(\theta, s, \varepsilon, X_s) - \mathbb{E}_\theta\nu(\theta, s, \varepsilon, X_s)))^{2p}e^{-pK(\theta)(t-s)}ds.$$

Therefore, this expectation is uniformly bounded on $t \geq 0, \varepsilon > 0$. □

Proof of Corollary 1 . We have $D(\theta, t, \varepsilon, X_t) = \int(\Phi(\beta, X_t - y) - \Phi(\beta, X_t - x_t(\alpha)))u_t^{\theta,\varepsilon}(dy)$. Similarly to the study of $T_2(t)$, a Taylor expansion of $\Phi(\beta, .)$ yields, using (40),

$$\Phi(\beta, X_t - y) - \Phi(\beta, X_t - x_t(\alpha)) = \frac{\partial \Phi}{\partial x}(\beta, 0))(x_t(\alpha) - y) + \rho_1(X_t, y) - \rho_2(X_t), \text{ with}$$

$$\rho_2(X_t) = \frac{1}{2}(X_t - x_t(\alpha))^3 \int_0^1 (1-u)^2 \frac{\partial^3 \Phi}{\partial x^3}(\beta, u(X_t - x_t(\alpha)))du. \tag{43}$$

Therefore, $D(\theta, t, \varepsilon, X_t) = \frac{\partial \Phi}{\partial x}(\beta, 0))(x_t(\alpha) - \mathbb{E}_\theta X_t) + \int \rho_1(X_t, y) u_t^{\theta,\varepsilon}(dy) - \rho_2(X_t)$. Using (41),

$$\mathbb{E}_\theta D(\theta, t, \varepsilon, X_t) = \frac{\partial \Phi}{\partial x}(\beta, 0))(x_t(\alpha) - \mathbb{E}_\theta X_t) - \mathbb{E}_\theta \rho_2(X_t).$$

By Theorem 2, $\mathbb{E}_\theta|\rho_2(X_t)| \lesssim \varepsilon^3 O(1)$. This yields (12).
Moreover, as for the upper bound of $T_2(t)$, $\mathbb{E}_\theta|\rho_1(X_t, \bar{X}_t)|^{2p} \lesssim \mathbb{E}_\theta|X_t - \overline{X}_t|^{6p} \lesssim \varepsilon^{6p}$.
By Theorem 2, uniformly on $t > 0$, $\mathbb{E}_\theta|\rho_2(X_t)|^{2p} \lesssim \mathbb{E}_\theta|X_t - x_t(\alpha)|^{6p} \leq \varepsilon^{6p} O(1)$.
Joining these two inequalities, we get (13). □

5.2. **Proofs of Section 3.** We start with two preliminary propositions useful for the inference.
Proof of Proposition 1.
Set $x_1(t) = x_t(\alpha), x_2(t) = \frac{\partial x_t}{\partial \alpha}(\alpha, t), x_3(t) = \frac{\partial^2 x_t}{\partial \alpha^2}(\alpha, t)$ and $\mathbf{x}(t) = [x_1(t)\, x_2(t)\, x_3(t)]'$. Then, $\mathbf{x}(t)$ is solution of the ordinary differential equation

$$d\mathbf{x}(t) = B(\mathbf{x}(t))dt, \quad B(\mathbf{x}(t)) = [B_1(\mathbf{x}(t))\ B_2(\mathbf{x}(t))\ B_3(\mathbf{x}(t))]' \tag{44}$$

where $B_1(\mathbf{x}) = V(\alpha, x_1), B_2(\mathbf{x}) = \frac{\partial V}{\partial \alpha}(\alpha, x_1) + \frac{\partial V}{\partial x}(\alpha, x_1)x_2$, and $B_3(\mathbf{x}) = \frac{\partial^2 V}{\partial \alpha^2}(\alpha, x_1) + [\frac{\partial V}{\partial x}(\alpha, x_1) + 2\frac{\partial^2 V}{\partial x \partial \alpha}(\alpha, x_1)]x_2 + \frac{\partial V}{\partial x}(\alpha, x_1)x_2^2 + \frac{\partial V}{\partial x}(\alpha, x_1)x_3$. We easily check that $B(\mathbf{x}^*) = 0$ for $\mathbf{x}^* = [x_1^*\, x_2^*\, x_3^*]'$ with, using (16) for $\ell(\alpha)$,

$$x_1^* = x^*(\alpha), \qquad x_2^* = \frac{1}{\ell(\alpha)} \frac{\partial V}{\partial \alpha}(\alpha, x_1^*), \tag{45}$$

$$x_3^* = \frac{1}{\ell(\alpha)} \left(\frac{\partial^2 V}{\partial \alpha^2}(\alpha, x_1^*) + [\frac{\partial V}{\partial x}(\alpha, x_1^*) + 2\frac{\partial^2 V}{\partial x \partial \alpha}(\alpha, x_1^*)]x_2^* + \frac{\partial V}{\partial x}(\alpha, x_1^*)x_2^*)^2 \right). \tag{46}$$

To check if this point is asymptotically stable, we compute $DB(\mathbf{x}^*) = [\frac{\partial B_i}{\partial x_j}(\mathbf{x}^*)]_{1 \leq i,j \leq 3}$. The matrix $DB(\mathbf{x})$ is triangular with diagonal elements equal to $-\ell(\alpha) < 0$. Thus, the eigenvalues of $DB(\mathbf{x}^*)$ are negative which implies that $\mathbf{x}^*$ is asymptotically stable for (44). Thus $\mathbf{x}(t)$ converges as $t \to +\infty$ to $\mathbf{x}^*$ with exponential rate $\exp(-\ell(\alpha)t)$ (see *e.g.* Hirsch and Smale, 1974). □

Note that $\alpha \to x^*(\alpha)$ is C^∞ on U_α. As Θ_α is compact, $\frac{dx^*}{d\alpha}(\alpha)$ and $\frac{d^2x^*}{d\alpha}(\alpha)$ are uniformly bounded on Θ_α as well as $x^*(\alpha)$.

Proof of Proposition 2.
Proof of (i): Consider the process (g_t) such that $dg_t = -\lambda g_t dt + dW_t,\ g_t(0) = 0$ with $\lambda > 0$. It is standard that $g_t = \exp(-\lambda t) \int_0^t \exp(\lambda s) dW_s$ and that (g_t) is a positive recurrent diffusion with invariant distribution $\mathcal{N}(0, 1/(2\lambda))$. By the ergodic theorem, $\frac{1}{T}\int_0^T g_s^2 ds$ converges a.s. to $(1/2\lambda)$. This implies, by the central limit theorem for martingales, that $\frac{1}{\sqrt{T}}\int_0^T g_s dW_s$ converges in distribution to $\mathcal{N}(0, (1/2\lambda))$. Moreover, one easily gets

$$\mathbb{E}(\frac{1}{T}\int_0^T g_s^2 ds) = \frac{1}{T}\int_0^T \mathbb{E} g_s^2 ds = \frac{1}{2\lambda} + o(1).$$

We can also compute

$$\mathbb{E}\frac{1}{T^2}\left(\int_0^T g_s^2 ds\right)^2 = \frac{1}{T^2}\int_{[0,T]^2} \mathbb{E}(g_s^2 g_t^2) ds dt.$$

Now, if $(X,Y) \sim \mathcal{N}_2(0,\Sigma)$ with $\Sigma = \begin{pmatrix} \sigma^2 & \alpha \\ \alpha & \tau^2 \end{pmatrix}$, then $\mathbb{E}(X^2Y^2) = \sigma^2\tau^2 + 2\alpha^2$. Applying this property to the centered Gaussian process (g_t), $\mathbb{E}(g_s^2 g_t^2) = 2\text{cov}^2(g_s,g_t) + \mathbb{E}g_s^2\mathbb{E}g_t^2$. Therefore,

$$\mathbb{E}\frac{1}{T^2}\left(\int_0^T g_s^2 ds\right)^2 = \left(\frac{1}{T}\int_0^T \mathbb{E}g_s^2 ds\right)^2 + C_T(\lambda). \tag{47}$$

where

$$C_T(\lambda) = \frac{2}{T^2}\int_{[0,T]^2} \text{cov}^2(g_s,g_t)dsdt = \frac{4}{T^2}\int_{0\le s\le t\le T} \text{cov}^2(g_s,g_t)dsdt. \tag{48}$$

For $s \le t$, $\text{cov}(g_s,g_t) = \int_0^s \exp -\lambda(t-u+s-u)du = \frac{1}{2\lambda}\left(\exp\left(\lambda(s-t) - \exp\left(\lambda(s+t)\right)\right.\right.$.
By elementary computations, we see that $C_T(\lambda) = \frac{1}{T}O(1)$ so that

$$\mathbb{E}\left(\frac{1}{T}\int_0^T g_s^2 ds - \frac{1}{T}\int_0^T \mathbb{E}g_s^2 ds\right)^2 \to 0.$$

With this direct calculus, we have obtained that $\frac{1}{T}\int_0^T g_s^2 ds \to_{\mathbb{L}^2} \frac{1}{2\lambda}$.

We rely on this approach to prove Proposition 2 for the process $g_t(\theta)$. Using (4) , (8) and (42), we have that under [H3], for $u \le t$,

$$A(\theta,t) - A(\theta,u) \le -K(\theta)(t-u). \tag{49}$$

Moreover, by (7), $g_t(\theta) = \int_0^t e^{A(\theta,t)-A(\theta,s)}dW_s$. Equations (47)-(48) hold for $g_t(\theta)$,

$$\mathbb{E}_\theta\frac{1}{T^2}\left(\int_0^T [g_s(\theta)]^2 ds\right)^2 = \left(\frac{1}{T}\int_0^T \mathbb{E}_\theta[g_s(\theta)]^2 ds\right)^2 + \tilde{C}_T(\theta) \quad \text{with}$$

$\tilde{C}_T(\theta) = \frac{4}{T^2}\int_{0\le s\le t\le T} \text{cov}_\theta^2(g_s(\theta),g_t(\theta))dsdt$. For $s \le t$, using (42) and (49)

$$\text{cov}_\theta(g_s(\theta),g_t(\theta)) = \int_0^s e^{A(\theta,t)+A(\theta,s)-2A(\theta,u)}du \le \int_0^s \exp\left[-(K(\theta)(t-u+s-u))\right]du.$$

Therefore $\tilde{C}_T(\theta) \le C_T(K(\theta))$. Finally, using (47)-(48),

$$\mathbb{E}_\theta\left(\frac{1}{T}\int_0^T [g_s(\theta)]^2 ds - \frac{1}{T}\int_0^T \mathbb{E}_\theta[g_s(\theta)]^2 ds\right)^2 \le C_T(K(\theta)) = \frac{1}{T}O(1).$$

Thus,

$$\frac{1}{T}\int_0^T [g_s(\theta)]^2 ds - \frac{1}{T}\int_0^T \mathbb{E}_\theta[g_s(\theta)]^2 ds \to_{\mathbb{L}^2} 0. \tag{50}$$

Now, the function $t \to \frac{\partial V}{\partial x}(\alpha, x_t(\alpha))$ is continuous. Under [H3]-[H4], as $t \to +\infty$, $x_t(\alpha) \to x^*(\alpha)$, and $\frac{\partial V}{\partial x}(\alpha, x_t(\alpha)) \to \frac{\partial V}{\partial x}(\alpha, x^*(\alpha)) = -\ell(\alpha) \le -K_V(\alpha) < 0$. Therefore,

$$\forall h > 0,\ \exists t_0 > 0,\ \forall t \ge t_0,\ -\ell(\alpha) - h < \frac{\partial V}{\partial x}(\alpha, x_t(\alpha)) < -\ell(\alpha) + h.$$

It follows, using (16), that for all t, s such that $t \ge s \ge t_0$,

$$-(\ell(\alpha,\beta) + h)(t-s) \le A(\theta,t) - A(\theta,s) \le -(\ell(\alpha,\beta) - h)(t-s). \tag{51}$$

Choose $h>0$ such that $\ell(\alpha,\beta)-h>0$. We have, using (7), $\mathbb{E}_\theta(g_t(\theta)^2)=e^{2A(\theta,t)}\int_0^t e^{-2A(\theta,s)}ds$. Hence, $\mathbb{E}_\theta(T^{-1}\int_0^T[g_t(\theta)]^2dt)=T_1+T_2+T_3$ where

$$\begin{aligned} T_1 &= \frac{1}{T}\int_{0\le s\le t\le t_0} e^{2(A(\theta,t)-A(\theta,s)}dsdt, \quad T_2=\frac{1}{T}\int_{0\le s\le t_0, t_0\le t\le T} e^{2(A(\theta,t)-A(\theta,s))}dsdt,\\ T_3 &= \frac{1}{T}\int_{t_0\le s\le t\le T} e^{2(A(\theta,t)-A(\theta,s))}dsdt. \end{aligned}$$

As T tends to infinity, $T_1=o(1)$. For T_2 we have,

$$T_2=\frac{1}{T}\int_0^{t_0} e^{-2(A(\theta,s)-A(\theta,t_0))}ds\times\int_{t_0}^T e^{2(A(\theta,t)-A(\theta,t_0))}dt.$$

Now, using (51),

$$\int_{t_0}^T e^{2(A(\theta,t)-A(\theta,t_0))}dt\le\int_{t_0}^T e^{-2(\ell(\alpha,\beta)-h)(t-t_0)}dt\le\frac{1}{2(\ell(\alpha,\beta)-h)}.$$

Therefore $0\le T_2\le\frac{1}{T}O(1)$ and $T_2\to 0$ as $T\to\infty$. Now, let us examine T_3:

$$T_3\le\frac{1}{T}\int_{t_0}^T ds\int_s^T e^{-2(\ell(\alpha,\beta)-h)(t-s)}dt=\frac{1}{2(\ell(\alpha,\beta)-h)T}\left(T-t_0-\frac{1-e^{-2(\ell(\alpha,\beta)-h)(T-t_0)}}{2(\ell(\alpha,\beta)-h)}\right).$$

Therefore, $\overline{\lim}_{T\to+\infty}T_3\le\frac{1}{2(\ell(\alpha,\beta)-h)}$. Analogously, $\underline{\lim}_{T\to+\infty}T_3\ge\frac{1}{2(\ell(\alpha,\beta)+h)}$.
Therefore, $T_3\to\frac{1}{2(\ell(\alpha,\beta)}$ so that

$$\mathbb{E}_\theta(\frac{1}{T}\int_0^T[g_t(\theta)]^2dt)\to\frac{1}{2\ell(\alpha,\beta)}\ \text{ as }\ T\to\infty.$$

Using (50), the first item is proved.

Proof of (ii): Let $Z_T=\int_0^T g_t(\theta)dt$. Using (7) and interchanging the order of integrations yields:

$$Z_T=\int_0^T g_t(\theta)dt=\int_0^T e^{A(\theta,t)}\int_0^t e^{-A(\theta,s)}dW_sdt=\int_0^T e^{-A(\theta,s)}dW_s\int_s^T e^{A(\theta,t)}dt.$$

Therefore, Z_T is centered and, using(49)

$$\begin{aligned}\mathbb{E}Z_T^2 &= \int_0^T e^{-2A(\theta,s)}ds\left(\int_s^T e^{A(\theta,t)}dt\right)^2=\int_0^T ds\left(\int_s^T e^{A(\theta,t)-A(\theta,s)}dt\right)^2\\ &\le \int_0^T ds\left(\int_s^T e^{-K(\theta)(t-s)}ds\right)^2\le\int_0^T ds\left(\frac{1-e^{-K(\theta)(T-s)}}{K(\theta)}\right)^2\le\frac{T}{K^2(\theta)}.\end{aligned}$$

Therefore, we find that $\mathbb{E}Z_T^2\lesssim T$ and $T^{-1}Z_T$ tends to 0 in probability as T tends to infinity.

Proof of (iii): As $\lim_{t\to+\infty}h(t)=0$, for all $h>0$, there exists $T_0>0$ such that for all $T\ge T_0$, $|h(t)|\le h$. So, we split

$$\frac{1}{\sqrt{T}}\int_0^T h(s)g_s(\theta)ds=\frac{1}{\sqrt{T}}\int_0^{T_0}h(s)g_s(\theta)ds+\frac{1}{\sqrt{T}}\int_{T_0}^T h(s)g_s(\theta)ds=o_P(1)+\frac{1}{\sqrt{T}}Z(T_0,T),\text{with}$$

$$\begin{aligned} Z(T_0,T) &= \int_{T_0}^T h(s)e^{A(\theta,s)}\left(\int_0^{T_0} e^{-A(\theta,u)}dW_u + \int_{T_0}^s e^{-A(\theta,u)}dW_u\right)ds \\ &= \int_0^{T_0} e^{-A(\theta,u)}dW_u \int_{T_0}^T h(s)e^{A(\theta,s)}ds + \int_{T_0}^T e^{-A(\theta,u)}dW_u \int_u^T h(s)e^{A(\theta,s)}ds = Z_{T,1}+Z_{T,2}. \end{aligned}$$

For the first term of $Z(T_0,T)$, $Z_{T,1}$, using (49) yields

$$|\int_{T_0}^T h(s)e^{A(\theta,s)}ds| \le h\int_{T_0}^T e^{-K(\theta)s}ds \le \frac{h}{K(\theta)}e^{-(K(\theta)T_0} = hO_P(1).$$

Hence $\mathbb{E}(Z_{T,1})^2 = h^2O(1)$. For the second term of $Z(T_0,T)$, we write

$$\begin{aligned} \mathbb{E}(Z_{T,2})^2 &= \int_{T_0}^T e^{-2A(\theta,u)}du\left(\int_u^T h(s)e^{A(\theta,s)}ds\right)^2 = \int_{T_0}^T du\left(\int_u^T h(s)e^{A(\theta,s)-A(\theta,u)}ds\right)^2 \\ &\le h^2\int_{T_0}^T du\left(\int_u^T e^{-K(\theta)(s-u)}ds\right)^2 \le (T-T_0)h^2\frac{1}{K(\theta)^2}. \end{aligned}$$

Therefore, for all $T\ge T_0$, $\frac{1}{T}\mathbb{E}(Z(T_0,T))^2 \lesssim \frac{h^2}{T}+h^2$. Hence, $\lim_{T\to+\infty}\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)h(s)ds = 0$. □

Proof of Theorem 4. Recall that $H(\theta,s,x) = V(\alpha,x) - \Phi(\beta, x - x_s(\alpha))$. Thus, using (11),

$$dX_s = \varepsilon dW_s + H(\theta,s,X_s)ds - D(\theta,s,\varepsilon,X_s)ds. \tag{52}$$

The derivatives of H with respect to the parameters are given by:

$$\begin{aligned} \frac{\partial H}{\partial\alpha}(\theta,s,X_s) &= \frac{\partial V}{\partial\alpha}(\alpha,X_s) + \frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))\frac{\partial x_s}{\partial\alpha}(\alpha,s), \\ \frac{\partial H}{\partial\beta}(\theta,s,X_s) &= -\frac{\partial\Phi}{\partial\beta}(\beta,X_s-x_s(\alpha)) \\ \frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s) &= \frac{\partial^2 V}{\partial\alpha^2}(\alpha,X_s) + \frac{\partial\Phi}{\partial x}(\beta,X_s-x_s(\alpha))\frac{\partial^2 x_s}{\partial\alpha^2}(\alpha,s) - \frac{\partial^2\Phi}{\partial x^2}(\beta,X_s-x_s(\alpha))\left(\frac{\partial x_s}{\partial\alpha}(\alpha,s)\right)^2 \\ \frac{\partial^2 H}{\partial\beta^2}(\theta,s,X_s) &= -\frac{\partial^2\Phi}{\partial\beta^2}(\beta,X_s-x_s(\alpha)),\ \frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s) = \frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,X_s-x_s(\alpha))\frac{\partial x_s}{\partial\alpha}(\alpha,s). \end{aligned}$$

Note that for the convergence in distribution stated in Theorem 4, it is enough to prove that

$$\frac{\varepsilon}{\sqrt{T}}\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta) = \ell(\alpha,\beta)\frac{\frac{\partial V}{\partial\alpha}(\alpha,x^*(\alpha))}{\ell(\alpha)}\frac{W_T}{\sqrt{T}} + o_P(1), \tag{53}$$

$$\frac{1}{\sqrt{T}}\frac{\partial\Lambda_{\varepsilon,T}}{\partial\beta}(\theta) = -\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0))\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)dW_s + o_P(1). \tag{54}$$

Indeed, the bracket of the two stochastic integrals above is equal, up to a constant, to $T^{-1}\int_0^T g_s(\theta)ds$ and tends to 0 as T tends to infinity by Proposition 5.
Moreover, as $T^{-1}\int_0^T[g_s(\theta)]^2ds$ tends to $[2\ell(\alpha,\beta)]^{-1}$, by the central limit theorem for martingales, $\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)dW_s$ converges in distribution to $\mathcal{N}(0,[2\ell(\alpha,\beta)]^{-1})$.
The proof of (53)-(54) relies on the following Lemma:

Lemma 3. *Let $F(\theta,s,x)$ a continuous function on $\Theta\times\mathbb{R}^+\times\mathbb{R}$, differentiable with respect to x and assume that there exist $C>0$ and a nonnegative integer c such that,*

$$\forall\theta\in\theta,\ \forall s\ge 0,\ |F(\theta,s,x)|\le C(1+|x|^c) \quad \textit{and } |\frac{\partial F}{\partial x}(\theta,s,x)|\le C(1+|x|^c). \tag{55}$$

Then, for $T \geq 1$, $\varepsilon \leq 1$, $D(\theta, s, \varepsilon, x)$ given in (11), the following holds.

(i) $\mathbb{E}\int_0^T (F(\theta, s, X_s) - F(\theta, s, x_s(\alpha)))^2 ds \leq C_1(\theta, F)\, T\, \varepsilon^2$.

(ii) $\mathbb{E}\left|\int_0^T F(\theta, s, X_s) D(\theta, s, \varepsilon, X_s) ds\right| \leq C_2(\theta, F)\, T\, \varepsilon^2$.

(iii) *If* $\int_0^{+\infty} |F(\theta, s, x_s(\alpha))| ds < +\infty$, *then* $\mathbb{E}\left|\int_0^T F(\theta, s, X_s) D(\theta, s, \varepsilon, X_s) ds\right| \leq C_3(\theta, F)(\varepsilon^2 + \varepsilon^3 T)$.

where the constants $C_i(\theta, F)$ only depend on F and θ.

Note that the functions $F(\theta, s, x) = H(\theta, s, x)$, $\frac{\partial H}{\partial \alpha}(\theta, s, x)$, $\frac{\partial^2 H}{\partial \alpha^2}(\theta, s, x)$ satisfy (55) under [S1] so that Lemma 3 holds for these functions.

We now start the proof of (53)-(54).

Derivative of the contrast with respect to α

Replacing dX_s by its expression, we get (see (11), (14), (15) and (52)):

$$\begin{aligned} \frac{\partial \Lambda_{\varepsilon,T}}{\partial \alpha}(\theta) &= \frac{1}{\varepsilon^2}\left(\int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s) dX_s - \int_0^T H(\theta, s, X_s) \frac{\partial H}{\partial \alpha}(\theta, s, X_s) ds\right), \\ &= \frac{1}{\varepsilon}\int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s) dW_s - \frac{1}{\varepsilon^2}\left(\int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s) D(\theta, s, \varepsilon, X_s) ds\right). \end{aligned} \tag{56}$$

Let us define

$$\frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha)) = \frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) + \frac{\partial \Phi}{\partial x}(\beta, 0) \frac{dx^*}{d\alpha}(\alpha). \tag{57}$$

Then

$$\frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) - \frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha)) = \frac{\partial V}{\partial \alpha}(\alpha, x_s(\alpha)) - \frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) + \frac{\partial \Phi}{\partial x}(\beta, 0)\left(\frac{\partial x_s}{\partial \alpha}(\alpha, s) - \frac{dx^*}{d\alpha}(\alpha)\right).$$

Therefore $\frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha))$ is the limit of $\frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha))$ as $s \to \infty$. Since we are in Case **(1)**, (17) yields using (16) that $\frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) = \ell(\alpha)\frac{dx^*}{d\alpha}(\alpha) \neq 0$ and

$$\frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha)) = \ell(\alpha, \beta)\frac{dx^*}{d\alpha}(\alpha) = \frac{\ell(\alpha, \beta)}{\ell(\alpha)}\frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha)) \neq 0.$$

Consequently, we can write

$$\begin{aligned} \frac{\varepsilon}{\sqrt{T}}\frac{\partial \Lambda_{\varepsilon,T}}{\partial \alpha}(\theta) &= \frac{W_T}{\sqrt{T}}\frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha)) \\ &\quad + \frac{1}{\sqrt{T}}\int_0^T \left(\frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) - \frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha))\right) dW_s \qquad (58) \\ &\quad + \frac{1}{\sqrt{T}}\int_0^T \left(\frac{\partial H}{\partial \alpha}(\theta, s, X_s) - \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha))\right) dW_s \qquad (59) \\ &\quad - \frac{1}{\varepsilon\sqrt{T}}\left(\int_0^T \frac{\partial H}{\partial \alpha}(\theta, s, X_s) D(\theta, s, \varepsilon, X_s) ds\right). \qquad (60) \end{aligned}$$

Using Proposition 1 and [S1], $\frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) - \frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha))$ converges exponentially fast to 0 so that $\int_0^{+\infty} \left(\frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) - \frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha))\right)^2 ds < +\infty$. Thus,

$$\int_0^T \left(\frac{\partial H}{\partial \alpha}(\theta, s, X_s) - \frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha))\right) dW_s \to_{T\to+\infty} \int_0^{+\infty} \left(\frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha)) - \frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha))\right) dW_s.$$

Therefore, (58) is $O_P(1/\sqrt{T})$ and tends to 0. By Lemma 3 (i),

$$\frac{1}{T}\mathbb{E}_\theta \int_0^T \left(\frac{\partial H}{\partial \alpha}(\theta, s, X_s) - \frac{\partial H}{\partial \alpha}(\theta, s, x_s(\alpha))\right)^2 ds \lesssim \varepsilon^2,$$

so that (59) is $O_P(\varepsilon)$. Lemma 3 (ii) yields

$$\frac{1}{\varepsilon\sqrt{T}}\mathbb{E}_\theta \int_0^T \left|\frac{\partial H}{\partial \alpha}(\theta, s, X_s) D(\theta, s, \varepsilon, X_s)\right| ds \lesssim \varepsilon\sqrt{T},$$

so that (60) is also $o_P(1)$ under the condition $\varepsilon\sqrt{T} \to 0$. So we find that

$$\frac{\varepsilon}{\sqrt{T}}\frac{\partial \Lambda_{\varepsilon,T}}{\partial \alpha}(\theta) = \frac{W_T}{\sqrt{T}}\frac{\partial H}{\partial \alpha}(\theta, x^*(\alpha)) + o_P(1) = \ell(\alpha, \beta)\frac{\frac{\partial V}{\partial \alpha}(\alpha, x^*(\alpha))}{\ell(\alpha)}\frac{W_T}{\sqrt{T}} + o_P(1), \tag{61}$$

which gives (53).

Derivative of the contrast with respect to β

We have:

$$\begin{aligned}
\frac{\partial \Lambda_{\varepsilon,T}}{\partial \beta}(\theta) &= \frac{1}{\varepsilon^2}\int_0^T \frac{\partial H}{\partial \beta}(\theta, s, X_s) dX_s - \frac{1}{\varepsilon^2}\int_0^T H(\theta, s, X_s)\frac{\partial H}{\partial \beta}(\theta, s, X_s) ds \\
&= -\frac{1}{\varepsilon}\int_0^T \frac{\partial \Phi}{\partial \beta}(\beta, X_s - x_s(\alpha)) dW_s + \frac{1}{\varepsilon^2}\int_0^T \frac{\partial \Phi}{\partial \beta}(\beta, X_s - x_s(\alpha)) D(\theta, s, \varepsilon, X_s) ds \qquad (62) \\
&:= T_1 + T_2. \qquad (63)
\end{aligned}$$

Since $x \to \frac{\partial \Phi}{\partial \beta}(\beta, x)$ is an odd function, $\frac{\partial \Phi}{\partial \beta}(\beta, 0) = 0$, so

$$\frac{\partial \Phi}{\partial \beta}(\beta, x) = x\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta, 0) + x^2 \int_0^1 (1-u)\frac{\partial^3 \Phi}{\partial \beta \partial x^2}(\beta, ux) du. \tag{64}$$

Replacing x by $X_s - x_s(\alpha)) = \varepsilon g_s(\theta) + \varepsilon^2 R_s^\varepsilon(\theta)$ yields that

$$T_1 = -\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta, 0)\left(\int_0^T g_s(\theta) dW_s\right) - T_{11},$$

with

$$T_{11} = \varepsilon \int_0^T R_s^\varepsilon(\theta) dW_s - \varepsilon^{-1}\int_0^T (X_s - x_s(\alpha))^2 \int_0^1 (1-u)\frac{\partial^3 \Phi}{\partial \beta \partial x^2}(\beta, u(X_s - x_s(\alpha))) du\, dW_s.$$

Thus,

$$\frac{1}{\sqrt{T}}T_1 = -\frac{\partial^2 \Phi}{\partial \beta \partial x}(\beta, 0)\left(\frac{1}{\sqrt{T}}\int_0^T g_s(\theta) dW_s\right) - \frac{1}{\sqrt{T}}T_{11}.$$

We have, by Theorem 3,

$$\mathbb{E}_\theta \int_0^T (R_s^\varepsilon(\theta))^2 ds \le 2\mathbb{E}_\theta \int_0^T (R_s^\varepsilon(\theta) - \mathbb{E}_\theta R_s^\varepsilon(\theta))^2 ds + 2\int_0^T (\mathbb{E}_\theta R_s^\varepsilon(\theta))^2 ds \lesssim TO(1),$$

where $O(1)$ does not depend on T and ε. This implies

$$\frac{1}{T}\mathbb{E}_\theta\left(\varepsilon\int_0^T R_s^\varepsilon(\theta)dW_s\right)^2=\frac{\varepsilon^2}{T}\mathbb{E}_\theta\int_0^T(R_s^\varepsilon(\theta))^2ds\lesssim\frac{\varepsilon^2}{T}\times T=\varepsilon^2.$$

Then, using [S1],

$$\frac{1}{T}\mathbb{E}_\theta\left(\varepsilon^{-1}\int_0^T(X_s-x_s(\alpha))^2\int_0^1(1-u)\frac{\partial^3\Phi}{\partial\beta\partial x^2}(\beta,u(X_s-x_s(\alpha)))du\,dW_s\right)^2\lesssim$$
$$\frac{1}{\varepsilon^2T}\mathbb{E}_\theta\left(\int_0^T(X_s-x_s(\alpha))^4(1+(X_s-x_s(\alpha))^{2c}ds\right)\lesssim\frac{1}{\varepsilon^2T}\times\varepsilon^4T=\varepsilon^2.$$

Therefore,

$$\frac{1}{\sqrt{T}}T_1=-\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\left(\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)dW_s\right)+O_P(\varepsilon).$$

For T_2, we have using (64),

$$\begin{aligned}T_2&=\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\frac{1}{\varepsilon^2}\int_0^T(X_s-x_s(\alpha))D(\theta,s,\varepsilon,X_s)ds\\&+\frac{1}{\varepsilon^2}\int_0^T(X_s-x_s(\alpha))^2\int_0^1(1-u)\frac{\partial^3\Phi}{\partial\beta\partial x^2}(\beta,u(X_s-x_s(\alpha)))duD(\theta,s,\varepsilon,X_s)ds.\end{aligned}$$

We split $D(\theta,s,\varepsilon,X_s)=\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)+D(\theta,s,\varepsilon,X_s)-\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)$ and use Corollary 1 and Theorem 2. The main term of $|T_2|/\sqrt{T}$ is $\frac{1}{\varepsilon^2\sqrt{T}}|\int_0^T(X_s-x_s(\alpha))\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)ds|$. Taking the expectation of this term yields

$$\mathbb{E}_\theta\frac{1}{\varepsilon^2\sqrt{T}}|\int_0^T(X_s-x_s(\alpha))\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)ds|\le$$
$$\frac{1}{\sqrt{T}}\varepsilon^{-2}\sup_{s,\varepsilon}|\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)|\int_0^T\mathbb{E}_\theta|X_s-x_s(\alpha)|ds\lesssim\frac{1}{\sqrt{T}}O(1)\times\varepsilon T=O(1)\times\varepsilon\sqrt{T}.$$

Finally, $T_2/\sqrt{T}=\varepsilon\sqrt{T}O_P(1)$.

Therefore,

$$\frac{1}{\sqrt{T}}\frac{\partial\Lambda_{\varepsilon,T}}{\partial\beta}(\theta)=-\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\left(\frac{1}{\sqrt{T}}\int_0^T g_\theta(s)dW_s\right)+o_P(1).$$

This yields (54). Hence the first part of Theorem 4, is proved.

It remains to study the limit of the matrix $D_{\varepsilon,T}^{(1)}\mathcal{J}_{\varepsilon,T}(\theta)D_{\varepsilon,T}^{(1)}=-\begin{pmatrix}\frac{\varepsilon^2}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta) & \frac{\varepsilon}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta)\\ \frac{\varepsilon}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta) & \frac{1}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta)\end{pmatrix}$.

We have:

$$\begin{aligned}\frac{\varepsilon^2}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta)&=\frac{1}{T}\int_0^T\frac{\partial^2H}{\partial\alpha^2}(\theta,s,X_s)dX_s-\frac{1}{T}\int_0^T H(\theta,s,X_s)\frac{\partial^2H}{\partial\alpha^2}(\theta,s,X_s)ds\\&-\frac{1}{T}\int_0^T(\frac{\partial H}{\partial\alpha}(\theta,s,X_s))^2ds=T_1+T_2+T_3\quad\text{with}\end{aligned}$$

$T_1 = \frac{\varepsilon}{T}\int_0^T \frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s)dW_s$; $T_2 = -\frac{1}{T}\int_0^T \frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s)D(\theta,s,\varepsilon,X_s)ds$; $T_3 = -\frac{1}{T}\int_0^T \left(\frac{\partial H}{\partial\alpha}(\theta,s,X_s)\right)^2 ds$.
For T_1, we write

$$T_1 = \frac{\varepsilon}{T}\int_0^T \frac{\partial^2 H}{\partial\alpha^2}(\theta,s,x_s(\alpha))dW_s + \frac{\varepsilon}{T}\int_0^T \left(\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,X_s) - \frac{\partial^2 H}{\partial\alpha^2}(\theta,s,x_s(\alpha))\right)dW_s.$$

Noting that $\frac{\partial^2\Phi}{\partial x^2}(\beta,x)$ is odd,

$$\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,x_s(\alpha)) = \frac{\partial^2 V}{\partial\alpha^2}(\alpha,x_s(\alpha)) + \frac{\partial\Phi}{\partial x}(\beta,0))\frac{\partial^2 x_s}{\partial\alpha^2}(\alpha,s).$$

This function is uniformly bounded thanks to Proposition 1. Therefore, using Lemma 3,

$$\mathbb{E}_\theta T_1^2 \lesssim \frac{\varepsilon^2}{T^2}\times(T+\varepsilon^2 T) = \frac{\varepsilon^2}{T}(1+\varepsilon^2) = o(1).$$

By Lemma 3, $\mathbb{E}_\theta(|T_2| \lesssim \frac{1}{T}\times\varepsilon^2 T = \varepsilon^2$.
For the last and main term T_3, we write (see (57)):

$$\begin{aligned} T_3 &= -\frac{1}{T}\int_0^T \left[\left(\frac{\partial H}{\partial\alpha}(\theta,s,X_s)\right)^2 - \left(\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right)^2\right]ds \\ &\quad -\frac{1}{T}\int_0^T \left[\left(\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right)^2 - \left(\frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha))\right)^2\right]ds - \left(\frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha))\right)^2. \end{aligned}$$

For the first term, we use Lemma 3 to prove that it is $o_P(1)$. For the second term, we use that $\left(\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right)^2$ converges to $\left(\frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha))\right)^2$ with exponential rate and this implies that this second term is $o(1)$. Hence T_3 tends to $-\left(\frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha))\right)^2$.
Joining these results, we have proved that $\frac{\varepsilon^2}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta)$ tends to $-\left(\frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha))\right)^2$.

Let us now study $\frac{1}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta)$. Using that $\frac{\partial^2 H}{\partial\beta^2}(\theta,s,X_s) = -\frac{\partial^2\Phi}{\partial\beta^2}(\beta,X_s-x_s(\alpha))$ yields

$$\begin{aligned} \frac{1}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta) &= -\frac{1}{\varepsilon T}\int_0^T \frac{\partial^2\Phi}{\partial\beta^2}(\beta,X_s-x_s(\alpha))dW_s + \frac{1}{\varepsilon^2 T}\int_0^T \frac{\partial^2\Phi}{\partial\beta^2}(\beta,X_s-x_s(\alpha))D(\theta,s,\varepsilon,X_s)ds \\ &\quad -\frac{1}{\varepsilon^2 T}\int_0^T \left(\frac{\partial\Phi}{\partial\beta}(\theta,X_s-x_s(\alpha))\right)^2 ds = S_1+S_2+S_3. \end{aligned}$$

The following relation is analogous to (64):

$$\frac{\partial^2\Phi}{\partial\beta^2}(\beta,x) = x\frac{\partial^3\Phi}{\partial\beta^2\partial x}(\beta,0) + x^2\int_0^1 (1-u)\frac{\partial^4\Phi}{\partial\beta^2\partial x^2}(\beta,ux)du. \tag{65}$$

Substituting x by $X_s - x_s(\alpha) = \varepsilon g_s(\theta) + \varepsilon^2 R^\varepsilon(s)$, we get that the main term of S_1 is

$$S_{11} = -\frac{1}{\varepsilon T}\int_0^T (X_s - x_s(\alpha))\frac{\partial^3\Phi}{\partial\beta^2\partial x}(\beta,0)dW_s = O_P(\frac{1}{\sqrt{T}}),$$

as, using Proposition 5,(i), $\mathbb{E}_\theta S_{11}^2 = O_P(1/T)$. For S_2, we split as previously
$D(\theta,s,\varepsilon,X_s) = \mathbb{E}_\theta D(\theta,s,\varepsilon,X_s) + D(\theta,s,\varepsilon,X_s) - \mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)$ and find that the main term of S_2 is

$$S_{22} = \frac{1}{\varepsilon^2 T}\int_0^T (X_s - x_s(\alpha))\frac{\partial^3\Phi}{\partial\beta^2\partial x}(\beta,0)\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)ds,$$

where $|\mathbb{E}_\theta S_{22}| \lesssim \varepsilon$ using Corollary 1 and Theorem 2.
The limit is obtained by S_3 whose main term is (see (64))

$$S_{33} = -\frac{1}{\varepsilon^2 T}\int_0^T \left((X_s - x_s(\alpha))\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\right)^2 ds = -\left(\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\right)^2 \frac{1}{T}\int_0^T g_s^2(\theta)ds + o_P(1).$$

Therefore, Proposition 5, (i) yields that S_{33} tends to $-\left(\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\right)^2/(2(\ell(\alpha,\beta))$. Joining these results, we get that the same holds for $\frac{1}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta)$.

It remains to study the off diagonal term $(\varepsilon/T)\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha\partial\beta}(\theta)$. We have

$$\begin{aligned}\frac{\varepsilon}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha\partial\beta}(\theta) &= \frac{1}{T}\int_0^T \frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s)dW_s - \frac{1}{\varepsilon T}\int_0^T D(\theta,s,\varepsilon,X_s)\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s)ds \\ &\quad -\frac{1}{\varepsilon T}\left(\int_0^T \frac{\partial H}{\partial\beta}(\theta,s,X_s)\frac{\partial H}{\partial\alpha}(\theta,s,X_s)ds\right) = T_1 + T_2 + T_3\end{aligned}$$

where

$$\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s) = \frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,X_s - x_s(\alpha))\frac{\partial x_s}{\partial\alpha}(\alpha,s),\ \frac{\partial H}{\partial\beta}(\theta,s,X_s) = -\frac{\partial\Phi}{\partial\beta}(\beta,X_s - x_s(\alpha)).$$

As before, the main term of T_1 is $\frac{1}{T}\int_0^T \frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,x_s(\alpha))dW_s = \frac{1}{T}\int_0^T \frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha,s)dW_s$. Therefore, since $\frac{\partial x_s}{\partial\alpha}(\alpha,s)$ is uniformly bounded, $\mathbb{E}_\theta T_1^2 = \frac{1}{T}O(1)$ and $T_1 = O_P(\frac{1}{\sqrt{T}})$.
For T_2, by Lemma 3, $|\mathbb{E}_\theta T_2| \lesssim \frac{1}{\varepsilon T}\varepsilon^2 T = \varepsilon$.

For T_3, we have, using (64),

$$T_3 = -\frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0)\frac{1}{T}\int_0^T g_s(\theta)\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))ds + o_P(1).$$

Now, setting $h(s) = \frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha)) - \frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha)$, we have

$$\frac{1}{T}\int_0^T g_s(\theta)\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))ds = \frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha))\frac{1}{T}\int_0^T g_s(\theta)ds + \frac{1}{T}\int_0^T g_s(\theta)h(s)ds.$$

Since $x_s(\alpha)\to x^*(\alpha)$, $h(s)\to 0$, Proposition 5 yields that both terms above converge to 0.
To conclude, we have obtained

$$\frac{\varepsilon}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta) = o_P(1).$$

The proof of Theorem 4 is now complete.□

Proof of Theorem 5 Let us set

$$h(\theta,s) = \frac{\partial V}{\partial\alpha}(\alpha,x_s(\alpha)) + \frac{\partial\Phi}{\partial x}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha,s) = \frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha)). \tag{66}$$

Here, for the convergence in distribution, it is enough to prove

$$\varepsilon\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta) = \int_0^T h(\theta,s)dW_s + o_P(1) \tag{67}$$

$$\frac{1}{\sqrt{T}}\frac{\partial\Lambda_{\varepsilon,T}}{\partial\beta}(\theta) = \frac{\partial^2\Phi}{\partial\beta\partial x}(\beta,0))\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)dW_s + o_P(1). \tag{68}$$

Indeed, the bracket of the two stochastic integrals is equal, up to a constant, to $\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)h(\theta,s)ds$. We are in Case **(2)**: using (17), it corresponds to $\frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha))=0$. Therefore, by Proposition 1, $h(\theta,s)$ converges exponentially fast to $\frac{\partial H}{\partial\alpha}(\theta,x^*(\alpha)=0$ and $\int_0^{+\infty}\left(\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right)^2 ds<+\infty$. Proposition 5 yields that $\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)h(\theta,s)ds$ tends to 0.

Let us prove (67). We now have (see (56)):

$$\begin{aligned}\varepsilon\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta) &= \int_0^T\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))dW_s+\int_0^T\left(\frac{\partial H}{\partial\alpha}(\theta,s,X_s)-\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right)dW_s\\ &\quad -\frac{1}{\varepsilon}\int_0^T\frac{\partial H}{\partial\alpha}(\theta,s,X_s)D(\theta,s,\varepsilon,X_s)ds=T_1+T_2+T_3. \qquad (69)\end{aligned}$$

Since $\mathbb{E}_\theta(T_1^2)<\infty$, $T_1\to\int_0^{+\infty}\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))dW_s$ as $T\to\infty$.
By Lemma 3, $\mathbb{E}_\theta(T_2^2)\lesssim\varepsilon^2T=o(1)$ under the condition $\varepsilon\sqrt{T}\to 0$.
As $\int_0^{+\infty}\left|\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right|ds<+\infty$, Lemma 3 (iii) yields that $\mathbb{E}|T_3|\lesssim\varepsilon+\varepsilon^2T=o(1)$.
This achieves the proof of (67).
The study of $\frac{1}{\sqrt{T}}\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\theta)$ is similar to its study in Theorem 4.The proof of (68) is complete.

Now we study the limit of the normalized matrix $D^{(2)}_{\varepsilon,T}\mathcal{J}_{\varepsilon,T}(\theta)D^{(2)}_{\varepsilon,T}=-\begin{pmatrix}\varepsilon^2\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta) & \frac{\varepsilon}{\sqrt{T}}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta)\\ \frac{\varepsilon}{\sqrt{T}}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta) & \frac{1}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta)\end{pmatrix}$.

We have

$$\begin{aligned}\varepsilon^2\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta) &= \varepsilon\int_0^T\frac{\partial^2H}{\partial\alpha^2}(\theta,s,X_s)dW_s-\int_0^T\frac{\partial^2H}{\partial\alpha^2}(\theta,s,X_s)D(\theta,s,\varepsilon,X_s)ds-\int_0^T\left(\frac{\partial H}{\partial\alpha}(\theta,s,X_s)\right)^2ds\\ &= T_1+T_2+T_3.\end{aligned}$$

For the first term, we write

$$T_1=\varepsilon\int_0^T\frac{\partial^2H}{\partial\alpha^2}(\theta,s,x_s(\alpha))dW_s+\varepsilon\int_0^T\left(\frac{\partial^2H}{\partial\alpha^2}(\theta,s,X_s)-\frac{\partial^2H}{\partial\alpha^2}(\theta,s,x_s(\alpha))\right)dW_s.$$

We have that $\frac{\partial^2H}{\partial\alpha^2}(\theta,s,x_s(\alpha))=\frac{\partial^2V}{\partial\alpha^2}(\alpha,x_s(\alpha))+\frac{\partial\Phi}{\partial x}(\beta,0))\frac{\partial^2x_s}{\partial\alpha^2}(\alpha,s)$, which is uniformly bounded on $\mathbb{R}^+$. Thus,

$$\mathbb{E}_\theta\left(\varepsilon\int_0^T\frac{\partial^2H}{\partial\alpha^2}(\theta,s,x_s(\alpha))dW_s\right)^2\lesssim\varepsilon^2T=o(1).$$

The second term of T_1 is ruled by Lemma 3 (i) and is $\varepsilon o_P(1)$. Next, $\mathbb{E}_\theta|T_2|\lesssim T\varepsilon^2$ by Lemma 3 (ii).

$$\mathbb{E}_\theta|\int_0^T\frac{\partial^2H}{\partial\alpha^2}(\theta,s,X_s)D(\theta,s,\varepsilon,X_s)ds|\lesssim\varepsilon^2T.$$

Finally, we can check, using Lemma 3 (i), that the main term of T_3 is, using (66), $\int_0^T h^2(\theta,s)ds$, where $h(\theta,s)$ converges exponentially fast to 0. Therefore,
$\int_0^T\left(\frac{\partial H}{\partial\alpha}(\theta,s,X_s)\right)^2ds\to\int_0^{+\infty}h^2(\theta,s)ds<+\infty$, so that

$$\varepsilon^2\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\theta)\to-\int_0^{+\infty}\left(\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))\right)^2ds.$$

The study of $\frac{1}{T}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta^2}(\theta)$ is the same as for Theorem 4. It remains to study

$$\begin{aligned}\frac{\varepsilon}{\sqrt{T}}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha\partial\beta}(\theta) &= \frac{1}{\sqrt{T}}\int_0^T\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s)dW_s - \frac{1}{\varepsilon\sqrt{T}}\int_0^T D(\theta,s,\varepsilon,X_s)\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,X_s)ds\\ &\quad -\frac{1}{\varepsilon\sqrt{T}}\left(\int_0^T\frac{\partial H}{\partial\beta}(\theta,s,X_s)\frac{\partial H}{\partial\alpha}(\theta,s,X_s)ds\right) = T_1+T_2+T_3.\end{aligned}$$

The proof is essentially analogous to the study of $\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha\partial\beta}(\theta)$ in the previous theorem. We point out the differences.

The main term of T_1 is $\frac{1}{\sqrt{T}}\int_0^T\frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha,s)dW_s$.

We are in Case **(2)** so that $\frac{\partial x_s}{\partial\alpha}(\alpha,s)$ converges exponentially fast to $\frac{dx^*}{d\alpha}(\alpha)=0$. Therefore, $\int_0^{+\infty}\left(\frac{\partial x_s}{\partial\alpha}(\alpha,s)\right)^2<+\infty$. Consequently, $T_1=\frac{1}{\sqrt{T}}O_P(1)$.

For T_2, the main term is $\frac{1}{\varepsilon\sqrt{T}}\int_0^T D(\theta,s,\varepsilon,X_s)\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,x_s(\alpha))ds$.

Using that $\frac{\partial^2 H}{\partial\alpha\partial\beta}(\theta,s,x_s(\alpha)) = \frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,0)\frac{\partial x_s}{\partial\alpha}(\alpha,s)$ is integrable, we get by Lemma 3 (iii), $\mathbb{E}_\theta[T_2|\lesssim\frac{1}{\varepsilon\sqrt{T}}(\varepsilon^2+\varepsilon^3T)=o(1)$.

It remains to study T_3. Using (64) and (66),

$$\begin{aligned}T_3 &= -\frac{1}{\varepsilon\sqrt{T}}\left(\int_0^T\frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,0)(X_s-x_s(\alpha))\frac{\partial H}{\partial\alpha}(\theta,s,x_s(\alpha))ds\right)ds+o_P(1)\\ &= -\frac{\partial^2\Phi}{\partial x\partial\beta}(\beta,0)\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)h(\theta,s)ds+o_P(1).\end{aligned}$$

As $h(\theta,s)\to 0$ as s tends to infinity, $\frac{1}{\sqrt{T}}\int_0^T g_s(\theta)h(\theta,s)ds=o_P(1)$ by Proposition 2.

We have obtained that $\frac{\varepsilon}{\sqrt{T}}\frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\beta\partial\alpha}(\theta)=o_P(1)$. So the proof of Theorem 5 is complete. □

Proof of Lemma 3.

Proof of (i) A Taylor expansion yields:

$$F(\theta,s,X_s)-F(\theta,s,x_s(\alpha)) = (X_s-x_s(\alpha))\int_0^1\frac{\partial F}{\partial x}(\alpha,x_s(\alpha)+u(X_s-x_s(\alpha)))du.$$

Hence

$$(F(\theta,s,X_s)-F(\theta,s,x_s(\alpha)))^2\le 3C^2\varepsilon^2\left(\frac{(X_s-x_s(\alpha))^2}{\varepsilon^2}(1+\sup_{s\ge0}|x_s(\alpha)|^{2c})+\varepsilon^{2c}\frac{(X_s-x_s(\alpha))^{2+2c}}{\varepsilon^{2+2c}}\right).$$

By Theorem 2, we get, using that $s\to x_s(\alpha)$ is uniformly bounded on $\mathbb{R}^+$ by $B(\alpha)$,

(70)

$$\mathbb{E}_\theta\int_0^T\left(F(\theta,sX_s)-F(\theta,s,x_s(\alpha))\right)^2ds\le 3C^2\varepsilon^2T\left(\delta(\alpha,1)(1+B^{2c}(\alpha))+\varepsilon^{2c}\delta(\alpha,1+c)\right)=C_1(\alpha,F)\varepsilon^2T.$$

This achieves the proof of (i).

Proof of (ii) For the second inequality, we split

$$\mathbb{E}_\theta \int_0^T F(\theta,s,X_s)D_s(\theta,s,\varepsilon,X_s)ds = A_1+A_2+A_3+A_4, \quad \text{with}$$

$$\begin{aligned}
A_1 &= \mathbb{E}_\theta \int_0^T F(\theta,s,x_s(\alpha))\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)ds,\\
A_2 &= \mathbb{E}_\theta \int_0^T F(\theta,s,x_s(\alpha))\left(D(\theta,s,\varepsilon,X_s)-\mathbb{E}D_\theta(\theta,s,\varepsilon,X_s)\right)ds,\\
A_3 &= \mathbb{E}_\theta \int_0^T \left(F(\theta,s,X_s)-F(\theta,s,x_s(\alpha))\right)\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)ds\\
A_4 &= \mathbb{E}_\theta \int_0^T \left(F(\theta,s,X_s)-F(\theta,s,x_s(\alpha))\right)\left(D(\theta,s,\varepsilon,X_s)-\mathbb{E}D_\theta(\theta,s,\varepsilon,X_s)\right)ds
\end{aligned}$$

Since $x_s(\alpha)$ is uniformly bounded by $B(\alpha)$, we get, using that $F(\theta,s,x)\le C(1+|x|^c)$, $|F(\theta,s,x_s(\alpha))|\le C(1+B^c(\alpha))=C(\alpha)$. Thererore, using Corollary 1

$$\begin{aligned}
|A_1| &\le \sup_{s\ge 0}|\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)|\int_0^T |F(\theta,s,x_s(\alpha))|\,ds \le \varepsilon^2 T\times\left[\sup_{s\ge0}|F(\theta,s,x_s(\alpha))|\,ds\right]\le C(\alpha)\varepsilon^2T\\
|A_2| &\le \int_0^T |F(\theta,s,x_s(\alpha))|\,\mathbb{E}_\theta\left|D(\theta,s,\varepsilon,X_s)-\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)\right|ds\\
&\le \varepsilon^3[\int_0^T |F(\theta,s,x_s(\alpha))|\left[\mathbb{E}_\theta\varepsilon^{-6}\left|D(\theta,s,\varepsilon,X_s)-\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)\right|^2\right]^{1/2}ds\lesssim\varepsilon^3T.
\end{aligned}$$

For A_3, we have using (i),

$$|A_3|\le\sup_{s\ge0}|\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)|\times\mathbb{E}_\theta\left[T\int_0^T|F(\theta,s,X_s)-F(\theta,s,x_s(\alpha)|^2\right]^{1/2}ds\lesssim\varepsilon^2\sqrt{T}\times(\varepsilon^2T)^{1/2}\lesssim\varepsilon^3T.$$

For A_4, we write:

$$\begin{aligned}
|A_4| &\le \int_0^T \mathbb{E}_\theta\left[|F(\theta,s,X_s)-F(\theta,s,x_s(\alpha))|\left|\left(D(\theta,s,\varepsilon,X_s)-\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)\right)\right|\right]ds\\
&\le \varepsilon^3\int_0^T ds\left[\mathbb{E}_\theta|F(\theta,s,X_s)-F(\theta,s,x_s(\alpha))|^2\,\mathbb{E}_\theta\left|\varepsilon^{-6}D(\theta,s,\varepsilon,X_s)-\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)\right|^2\right]^{1/2}.
\end{aligned}$$

We have, using (i), $\mathbb{E}_\theta|F(\theta,s,X_s)-F(\theta,s,x_s(\alpha))|^2\lesssim\left[\mathbb{E}_\theta(X_s-x_s(\alpha))^2\right]^{1/2}\le\varepsilon$.
Therefore, by Theorem 2 and Corollary 1 , $|A_4|\lesssim\varepsilon^4T$.
Finally, joining these inequalities yield (ii).

Proof of (iii) Since $\int_0^\infty|F(\theta,s,x_s(\alpha))|\,ds<\infty$, we can bound differently A_1 and A_2.

$$|A_1|\le\sup_{s\ge0}|\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)|\int_0^T|F(\theta,s,x_s(\alpha))|\,ds\lesssim\varepsilon^2.$$

Analogously, for A_2,

$$\begin{aligned}
|A_2| &\le [\int_0^T |F(\theta,s,x_s(\alpha))|\,\mathbb{E}_\theta\left|D(\theta,s,\varepsilon,X_s)-\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)\right|ds\\
&\le \varepsilon^3\sup_{s\ge0}\left(\mathbb{E}_\theta[\varepsilon^{-6}\left|D(\theta,s,\varepsilon,X_s)-\mathbb{E}_\theta D(\theta,s,\varepsilon,X_s)\right|^2]\right)^{1/2}[\int_0^{+\infty}|F(\theta,s,x_s(\alpha))|\,ds\lesssim\varepsilon^3.
\end{aligned}$$

The terms A_3, A_4 are bounded as previously. Thus $|A_1 +_2 +A_3 + A_4| \lesssim \varepsilon^2 + \varepsilon^3 T$.
It remains to look at the three functions $H(\theta,s,x)$, $\frac{\partial H}{\partial\alpha}(\theta,s,x)$, $\frac{\partial^2 H}{\partial\alpha^2}(\theta,s,x)$. Using [S1]-[S2], as $B = \sup_{\alpha,t}|x_t(\alpha)| < +\infty$, we easily check (55) for $H(\theta,s,x)$. By [S2] and Proposition 1, we also have $\sup_{\alpha,t}|\frac{\partial x_t}{\partial\alpha}(\alpha,t)| < +\infty$ and $\sup_{\alpha,t}|\frac{\partial^2 x_t}{\partial\alpha^2}(\alpha,t)| < +\infty$. Therefore, we can check that (55) holds for the two other functions. □

Proof of Lemma 2 We have to study under $\mathbb{P}_{\theta_0}$:

$$\begin{aligned}\varepsilon^2\Lambda_{\varepsilon,T}(\alpha,\beta) &= \int_0^T H(\theta,s,X_s)(V(\alpha_0,X_s)ds - b(\theta_0,s,X_s))ds + \varepsilon dW_s - \frac{1}{2}\int_0^T H^2(\theta,s,X_s)ds \\ &= -\frac{1}{2}\int_0^T (H(\theta,s,X_s) - V(\alpha_0,X_s))^2 ds + \frac{1}{2}\int_0^T V^2(\alpha_0,X_s)ds + \sum_{i=1}^4 T_i\end{aligned}$$

where, using (15),(11)

$$\begin{aligned}T_1 &= \varepsilon\int_0^T H(\theta,s,X_s)dW_s;\ \ T_2 = \int_0^T H(\theta,s,X_s)D(\theta_0,s,\varepsilon,X_s)ds \\ T_3 &= -\int_0^T (H(\theta,s,X_s) - H(\theta,s,x_s(\alpha_0)))\Phi(\beta_0, X_s - x_s(\alpha_0))ds \\ T_4 &= -\int_0^T H(\theta,s,x_s(\alpha_0))\Phi(\beta_0, X_s - x_s(\alpha_0))ds.\end{aligned}$$

Let us consider the first term of $\varepsilon^2\Lambda_{\varepsilon,T}(\alpha,\beta)$. It satisfisfies, using Lemma 3 (i) that , under the condition $\varepsilon\sqrt{T}\to 0$,
$\int_0^T (H(\theta,s,X_s) - V(\alpha_0,X_s))^2 ds = \int_0^T (H(\theta,s,x_s(\alpha_0)) - V(\alpha_0,x_s(\alpha_0)))^2 ds + o_P(1)$.
Now, define the limit of its integrand term as $s\to\infty$,

$$h^*(\alpha,\alpha_0,\beta) = V(\alpha,x^*(\alpha_0)) - \Phi(\beta, x^*(\alpha_0) - x^*(\alpha)). \tag{71}$$

The two cases pointed out in Section 3.3 occur here.
Case **(1)**: For all β, $h^*(\alpha,\alpha_0,\beta)\neq 0$, and $\frac{1}{T}\int_0^T (H(\theta,s,x_s(\alpha_0)) - V(\alpha_0,x_s(\alpha_0)))^2 ds \to (h^*(\alpha,\alpha_0,\beta))^2$.
Case **(2)**: For all β, $h^*(\alpha,\alpha_0,\beta) = 0$, and $\int_0^\infty (H(\theta,s,x_s(\alpha_0)) - V(\alpha_0,x_s(\alpha_0)))^2 ds < \infty$.

The second term satisfies, in both cases $\int_0^T V^2(\alpha_0,X_s)ds = \int_0^T V^2(\alpha_0,x_s(\alpha_0)))ds + o_P(1)$ and this last integral converges, as $T\to\infty$ to $\int_0^\infty V^2(\alpha_0,x_s(\alpha_0)))ds < \infty$.

Consider now the remainder terms T_i of $\varepsilon^2\Lambda_{\varepsilon,T}(\alpha,\beta)$.
We have $\mathbb{E}_{\theta_0}T_1^2 = \varepsilon^2\mathbb{E}_{\theta_0}\int_0^T [V(\alpha,X_s) - \Phi(\beta,X_s - x_s(\alpha))]^2 ds$. Using Lemma 3 (i) and similar tools detailed in the proof yields that $\mathbb{E}_{\theta_0}T_1^2 \lesssim \varepsilon^2 T$. Therefore, under the condition $\varepsilon\sqrt{T} = o(1)$ we find that $T_1 = o_P(1), T_2 = o_P(1)$. For T_3, applying Lemma 3 (i) yields that
$\mathbb{E}_{\theta_0}\int_0^T [H(\theta,s,X_s) - H(\theta,s,x_s(\alpha_0)))\Phi(\beta_0,X_s - x_s(\alpha_0)])^2 ds \lesssim T\varepsilon^2$ and $\mathbb{E}_{\theta_0}|T_3| = \varepsilon\sqrt{T} = o_P(1)$.
For T_4, using Theorem 3, $\Phi(\beta_0, X_s - x_s(\alpha_0)) = \frac{\partial\Phi}{\partial x}(\beta,0)(\varepsilon g_s(\theta_0) + \varepsilon^2 R_s^\varepsilon(\theta_0))$.
Therefore $T_4 = \varepsilon\dfrac{\partial\Phi}{\partial x}(\beta,0)\displaystyle\int_0^T H(\theta,s,x_s(\alpha_0))g_s(\theta_0)ds + o_P(1)$.
The limit, as $s\to\infty$ of $H(\theta,s,x_s(\alpha_0))$ is $h^*(\alpha_0,\alpha,\beta)$ defined in (71). Therefore, we have to study $\frac{T_4}{T}$ in Case **(1)** and T_4 in Case **(2)**. We have

$$T_4 = \varepsilon\frac{\partial\Phi}{\partial x}(\beta,0)\left(h^*(\alpha,\alpha_0,\beta)\int_0^T g_s(\theta_0)ds + \int_0^T (H(\theta,s,x_s(\alpha_0)) - h^*(\alpha,\alpha_0,\beta))g_s(\theta_0)ds\right).$$

Therefore, in Case **(1)**, applying Proposition 5 (ii) and (iii) yields that $\frac{T_4}{T} = \varepsilon o_P(1)$.
In Case **(2)**, for all β, $h^*(\alpha_0, \alpha, \beta) = 0$ and Proposition 5 (iii) yields that $T_4 = \varepsilon\sqrt{T}o_P(1) = o_P(1)$.

Consider now $\varepsilon^2\Lambda_{\varepsilon,T}(\alpha_0, \beta)$. Noting that $h^*(\alpha_0, \alpha_0, \beta) = 0$, we have that $T_4 = o_P(1)$. Using that $\int_0^T [\Phi(\beta, X_s - x_s(\alpha))]^2 ds = O_P(\varepsilon^2 T)$, we get

$$\begin{aligned}\varepsilon^2\Lambda_{\varepsilon,T}(\alpha_0, \beta) &= -\frac{1}{2}\int_0^T [\Phi(\beta, X_s - x_s(\alpha))]^2 ds + \frac{1}{2}\int_0^T V^2(\alpha_0, X_s)ds + o_P(1)\\ &= \frac{1}{2}\int_0^T V^2(\alpha_0, X_s)ds + o_P(1),\end{aligned}$$

Joining these results yields that, using (71),
Case **(1)**: $\frac{\varepsilon^2}{T}(\Lambda_{\varepsilon,T}(\alpha, \beta) - \Lambda_{\varepsilon,T}(\alpha_0, \beta)) \to -\frac{1}{2}(h^*(\alpha, \alpha_0, \beta))^2 = \Lambda_1^{(1)}(\alpha, \alpha_0, \beta)$.

Case **(2)**: Set $\Lambda_1^{(2)}(\alpha, \alpha_0, \beta) = -\frac{1}{2}\int_0^{+\infty}[V(\alpha, x_s(\alpha_0)) - V(\alpha_0, x_s(\alpha_0)) - \Phi(\beta, x_s(\alpha_0) - x_s(\alpha))]^2 ds$.
Then, $\varepsilon^2(\Lambda_{\varepsilon,T}(\alpha, \beta) - \Lambda_{\varepsilon,T}(\alpha_0, \beta)) \to \Lambda_1^{(2)}(\alpha, \alpha_0, \beta)$.

The uniformity of the convergence is obtained using that $\Theta_\alpha, \Theta_\beta$ are compact sets, Assumptions [S1], [S2] and Remarks 3, 5.

Finally, it remains to study $\frac{1}{T}(\Lambda_{\varepsilon,T}(\alpha_0, \beta) - \Lambda_{\varepsilon,T}(\alpha_0, \beta_0))$.

$$\begin{aligned}\varepsilon^2\Lambda_{\varepsilon,T}(\alpha_0, \beta) &= \int_0^T H(\alpha_0, \beta, s, X_s)[(H(\theta_0, s, X_s) - D(\theta_0, s, \varepsilon, X_s))ds + \varepsilon dW_s]\\ &- \frac{1}{2}\int_0^T H^2(\alpha_0, \beta, s, X_s)ds = -\frac{1}{2}\int_0^T (H(\alpha_0, \beta, s, X_s) - H(\theta_0, s, X_s))^2 ds\\ &+ \varepsilon\int_0^T H(\alpha_0, \beta, s, X_s)dW_s - \int_0^T H(\alpha_0, \beta, s, X_s)D(\theta_0, s, \varepsilon, X_s)ds.\end{aligned}$$

Using that $H(\alpha_0, \beta, s, X_s) - H(\theta_0, s, X_s) = -(\Phi(\beta, X_s - x_s(\alpha_0)) - \Phi(\beta_0, X_s - x_s(\alpha_0)))$ and

$$\Phi(\beta, X_s - x_s(\alpha_0)) = \frac{\partial\Phi}{\partial x}(\beta, 0)(X_s - x_s(\alpha_0)) + O_P(\varepsilon) = \varepsilon\frac{\partial\Phi}{\partial x}(\beta, 0)g_s(\theta_0) + \varepsilon^2 O_P(1).$$

Therefore,

$$\frac{1}{T}(\Lambda_{\varepsilon,T}(\alpha_0, \beta) - \Lambda_{\varepsilon,T}(\alpha_0, \beta_0)) = -\frac{1}{2T}\int_0^T (\frac{\partial\Phi}{\partial x}(\beta, 0) - \frac{\partial\Phi}{\partial x}(\beta_0, 0))^2 g_s^2(\theta_0)ds + T_1 + T_2 + \varepsilon O_P(1),$$

$T_1 = \frac{1}{\varepsilon T}\int_0^T (H(\alpha_0, \beta, s, X_s) - H(\theta_0, s, X_s))dW_s$, $T_2 = -\frac{1}{\varepsilon^2 T}\int_0^T (H(\alpha_0, \beta, s, X_s) - H(\theta_0, s, X_s))D(\theta_0, s, \varepsilon, X_s)ds$.
For T_1, we have, using Theorem 2,
$\mathbb{E}_{\theta_0} T_1^2 = \frac{1}{\varepsilon^2 T^2}\mathbb{E}_{\theta_0}\int_0^T (H(\alpha_0, \beta, s, X_s) - H(\theta_0, s, X_s))^2 ds \lesssim \frac{1}{\varepsilon^2 T^2} T \sup \mathbb{E}_{\theta_0}((X_s - x_s(\alpha_0))^2) \lesssim \frac{1}{T}$.
Therefore $T_1 = o_P(1)$.
For T_2, set $F(X_s) = H(\alpha_0, \beta, s, X_s) - H(\theta_0, s, X_s)$. Then, splitting $D(\theta_0, s, \varepsilon, X_s)$ as in the proof of Lemma 3,
$\int_0^T F(X_s)D(\theta_0, s, \varepsilon, X_s)ds = \int_0^T F(X_s)\mathbb{E}_{\theta_0} D(\theta_0, s, \varepsilon, X_s) + \int_0^T F(X_s)(D(\theta_0, s, \varepsilon, X_s) - \mathbb{E}_{\theta_0} D(\theta_0, s, \varepsilon, X_s))ds$.
Using that $\mathbb{E}_{\theta_0}|\int_0^T |F(X_s)|ds \le \sqrt{T}(\mathbb{E}_{\theta_0}\int_0^T F^2(X_s)ds)^{1/2} \le \varepsilon T$, we get
$\mathbb{E}_{\theta_0}|\int_0^T F(X_s)\mathbb{E}_{\theta_0} D(\theta_0, s, \varepsilon, X_s)ds| \le \sup_s |\mathbb{E}_{\theta_0} D(\theta_0, s, \varepsilon, X_s)|\ \mathbb{E}_{\theta_0}|\int_0^T |F(X_s)|ds \lesssim \varepsilon^3 T$.

Now, $\mathbb{E}_{\theta_0}|F(X_s)(D(\theta_0,s,\varepsilon,X_s)-\mathbb{E}_{\theta_0}D(\theta_0,s,\varepsilon,X_s))| \leq \varepsilon^3[\mathbb{E}_{\theta_0}(X_s-x_s(\alpha_0)^2]^{1/2}O(1)$.
Hence, $\mathbb{E}_{\theta_0}|\int_0^T F(X_s)(D(\theta_0,s,\varepsilon,X_s)-\mathbb{E}_{\theta_0}D(\theta_0,s,\varepsilon,X_s))ds| \leq \varepsilon^4 T$.

These two inequalities yield that $T_2 = o_P(1)$ and finally, as $T\to\infty$,

$$\begin{aligned}\frac{1}{T}(\Lambda_{\varepsilon,T}(\alpha_0,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta_0)) &= -\frac{1}{2T}\int_0^T [\frac{\partial\Phi}{\partial x}(\beta,0)-\frac{\partial\Phi}{\partial x}(\beta_0,0)]^2 g_s^2(\theta_0)ds + o_P(1)\\ &\to -\frac{1}{2\ell(\alpha_0,\beta_0)}[\frac{\partial\Phi}{\partial x}(\beta,0)-\frac{\partial\Phi}{\partial x}(\beta_0,0)]^2 = \Lambda_2(\alpha_0,\beta,\beta_0).\end{aligned}$$

Moreover, we can prove that this convergence is uniform with respect to $\beta\in\Theta_\beta$. □

Proof of Theorem 6. We just give here a sketch of the proof. To get (i), we prove the three steps (1)-(3) of Gloter and Sorensen (2009), Section 4.4.1, that we have recalled at the beginning of Section 3.4.

Proof of (1). In Case **(1)**, the fact that $(\varepsilon^2/T)(\Lambda_{\varepsilon,T}(\alpha,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta)) \to_{\mathbb{P}_{\theta_0}} \Lambda_1^{(1)}(\alpha,\alpha_0,\beta)$, uniformly with respect to (α,β) where $(\alpha,\beta)\to\Lambda_1^{1)}(\alpha,\alpha_0,\beta)$ is continuous, <0, and $=0$ iff $\alpha=\alpha_0$ implies the consistency of $\hat\alpha_{\varepsilon,T}$.

Analogously, in Case **(2)**, the fact that $\varepsilon^2(\Lambda_{\varepsilon,T}(\alpha,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta)) \to_{\mathbb{P}_{\theta_0}} \Lambda_1^{(2)}(\alpha,\alpha_0,\beta)$, uniformly with respect to (α,β) implies the consistency of $\hat\alpha_{\varepsilon,T}$.

Proof of (2). By (1), $\hat\alpha_{\varepsilon,T}$ is consistent thus $\mathbb{P}_{\theta_0}(\hat\alpha_{\varepsilon,T}\in\Theta_\alpha)\to 1$ as ε tends to 0. On the set $(\hat\alpha_{\varepsilon,T}\in\Theta_\alpha)$, we have:

$$\begin{aligned}0 &= \frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\hat\alpha_{\varepsilon,T},\hat\beta_{\varepsilon,T}) = V_{\varepsilon,T}+(\hat\alpha_{\varepsilon,T}-\alpha_0)N_{\varepsilon,T}, \quad \text{where}\\ V_{\varepsilon,T} &= \frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\alpha_0,\hat\beta_{\varepsilon,T)}, \quad N_{\varepsilon,T}=\int_0^1 \frac{\partial^2\Lambda_{\varepsilon,T}}{\partial\alpha^2}(\alpha_0+t(\hat\alpha_{\varepsilon,T}-\alpha_0),\hat\beta_{\varepsilon,T})dt.\end{aligned}$$

Thus,

$$\sqrt{T}\varepsilon^{-1}(\hat\alpha_{\varepsilon,T}-\alpha_0) = -\frac{(\varepsilon/\sqrt{T})V_{\varepsilon,T}}{(\varepsilon^2 T)N_{\varepsilon,T}} \text{ for Case } \mathbf{(1)}, \quad \varepsilon^{-1}(\hat\alpha_{\varepsilon,T}-\alpha_0) = -\frac{\varepsilon V_{\varepsilon,T}}{\varepsilon^2 N_{\varepsilon,T}} \text{ for Case } \mathbf{(2)}.$$

We must prove that $(\varepsilon/\sqrt{T})V_{\varepsilon,T}$ and $(\varepsilon^2/T)N_{\varepsilon,T}$ for Case **(1)**, $\varepsilon V_{\varepsilon,T}$ and $\varepsilon^2 N_{\varepsilon,T}$ for Case **(2)**, are tight under $\mathbb{P}_{\theta_0}$. This can be done using the same tools as in Theorems 4 and 5, and using the assumption that $\frac{\partial\Phi}{\partial x}(\beta,0)$ is uniformly bounded on Θ_β and that $\hat\beta_{\varepsilon,T}\in\Theta_\beta$.

Proof of (3). To obtain the consistency of $\hat\beta_{\varepsilon,T}$, it is enough to prove that:

$$\frac{1}{T}(\Lambda_{\varepsilon,T}(\hat\alpha_{\varepsilon,T},\beta)-\Lambda_{\varepsilon,T}(\hat\alpha_{\varepsilon,T},\beta_0)) \to \Lambda_2(\alpha_0,\beta,\beta_0) \tag{72}$$

uniformly in β.

Consider first Case **(1)**. Using (53), we have, setting $\alpha_u=\alpha_0+u(\hat\alpha_{\varepsilon,T}-\alpha_0)$,

$$\begin{aligned}\Lambda_{\varepsilon,T}(\hat\alpha_{\varepsilon,T},\beta)-\Lambda_{\varepsilon,T}(\hat\alpha_{\varepsilon,T},\beta_0) &= (\Lambda_{\varepsilon,T}(\alpha_0,\beta)-\Lambda_{\varepsilon,T}(\alpha_0,\beta_0))+\frac{\sqrt{T}}{\varepsilon}(\hat\alpha_{\varepsilon,T}-\alpha_0)\frac{\varepsilon}{\sqrt{T}}R(\varepsilon,\theta,T),\\ \text{with} \quad R(\varepsilon,\theta,T) &= \int_0^1 (\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\alpha_u,\beta)-\frac{\partial\Lambda_{\varepsilon,T}}{\partial\alpha}(\alpha_u,\beta_0))du\\ &= \frac{\sqrt{T}}{\varepsilon}\left(\int_0^1 \frac{(\ell(\alpha_u,\beta)-\ell(\alpha_u,\beta_0))}{\ell(\alpha_u)}\frac{\partial V}{\partial\alpha}(\alpha_u,x^*(\alpha_u))du\,\frac{W_T}{\sqrt{T}}+o_P(1)\right)\end{aligned}$$

Now, since $\hat{\alpha}_{\varepsilon,T}$ is consistent, the integral term converges to a constant $C(\theta_0,\beta)$ which is bounded. Therefore $\frac{1}{T^{3/2}}\varepsilon R(\varepsilon,\theta,T) = \frac{W_T}{T^{3/2}}C(\theta_0,\beta) + \frac{1}{T}o_P(1) = o_P(1)$.
Using now the tightness of $\varepsilon^{-1}\sqrt{T}(\hat{\alpha}_{\varepsilon,T} - \alpha_0)$ yields (72). The uniformity in β follows from the continuity of $\theta \to \ell(\theta)$.
For Case **(2)**, we use (66)-(67) and

$$R(\varepsilon,\theta,T) = \frac{1}{\varepsilon}(\int_0^T \int_0^1 (h(\alpha_u,\beta,s) - h(\alpha_u,\beta_0,s)du)dW_s + o_P(1))$$

Therefore $\frac{1}{T}\varepsilon R(\varepsilon,\theta,T) = \frac{1}{T}\int_0^T dW_s(\int_0^1 (h(\alpha_u,\beta,s) - h(\alpha_u,\beta_0,s)du) + \frac{1}{T}o_P(1)$ Since $\hat{\alpha}_{\varepsilon,T}$ is consistent, the integral term converges to a function $F(\theta_0,\beta,s)$ which is bounded uniformly in s. Hence $\mathbb{E}_{\theta_0}(\frac{1}{T}\varepsilon R(\varepsilon,\theta,T))^2 = \frac{1}{T^2}\int_0^T ds\mathbb{E}_{\theta_0}(\int_0^1 (h(\alpha_u,\beta,s) - h(\alpha_u,\beta_0,s)du)^2) \lesssim \frac{1}{T}O(1)$. Hence, we get that (72) also holds in Case **(2)**.
Under the identifiability asssumption for β, we get that in both cases the consistency of $\hat{\beta}_{\varepsilon,T}$.

The proof of the asymptotic normality follows, by standard tools from (i) and Theorems 4 and 5. □

References

[1] Altmeyer, R. and Reiss, M. (2020). Nonparametric estimation for linear SPDEs from local measurements. *Annals of Applied Probabilty* to appear.

[2] Amorino C. and Gloter, A. (2020). Contrast function estimation for the drift parameter of ergodic jump diffusion process. *Scand. J. Statist.* to appear.

[3] Ball, F. and Sirl D. (2020) Stochastic *SIR* in Structured Populations, in *Stochastic Epidemic Models with inference*, **Part II**, 123-240. Britton,T. and Pardoux, E., Editors. Lecture Notes in Mathematics 2255, Mathematical Biosciences Subseries. Springer.

[4] Baladron, J., Fasoli, D., Faugeras, O. and Touboul, J. (2012). Mean field description and propagation of chaos in networks of Hodgkin-Huxley and FitzHugh-Nagumo neurons. *Journal of Mathematical Neuroscience* **2:10**, 1-50.

[5] Benachour, S., Roynette, B., Talay, D. and Vallois, P. (1998a). Nonlinear self-stabilizing processes - I Existence, invariant probability, propagation of chaos. *Stoch. Proc. Appl.* **75**, 173-201.

[6] Benachour, S., Roynette, B. and Vallois, P. (1998b). Nonlinear self-stabilizing processes - II Convergence to invariant probability. *Stoch. Proc. Appl.* **75**, 203-224.

[7] Benedetto, D., Caglioti, E. and Pulverenti, M. (1997). A kinetic equation for granular media. *Mathematical Modelling and Numerical Analysis*, **31** (5), 615-641.

[8] Britton,T. and Pardoux, E., Editors (2020) *Stochastic Epidemic Models with Inference*. Lecture Notes in Mathematics 2255, Mathematical Biosciences Subseries. Springer.

[9] Cialenco, I. (2018). Statistical Inference for SPDEs: an Overview. *Statistical Inference for Stochastic Processes*, **21**, 309-329.

[10] Della Maestra, L. and Hoffmann, M. (2020). Nonparametric estimation for interacting particle systems: McKean-Vlasov models. Preprint arXiv:2011.03762 [math.ST].

[11] Comte, F. and Genon-Catalot, V. (2020). Non parametric drift estimation for *i.i.d.* paths of stochastic differential equations. *The Annals of Statistics* **48**, 3336-3365.

[12] Comte, F., Genon-Catalot, V. and Rozenholc, Y. (2007). Penalized nonparametric mean square estimation of the coefficients of diffusion processes. *Bernoulli*, **13**, 514-543.

[13] Dalalyan, A. (2005). Sharp adaptive estimation of the drift function for ergodic diffusions. *The Annals of Statistics*, **33**, 2507-2528.

[14] Dalalyan, A. and Reiss, M. (2007). Asymptotic statistical equivalence for ergodic diffusions: the multidimensional case. *Probab. Theory Relat. Fields*, **137**, 25-47.

[15] Delattre M., Genon-Catalot V. and Samson A. (2013). Maximum likelihood estimation for stochastic differential equations with random effects. *Scandinavian Journal of Statistics*, **40**, 322-343.

[16] Delattre M., Genon-Catalot V. and Larédo, C. (2018). Parametric inference for discrete observations of diffusion processes with mixed effects. *Stochastic processes and their Applications*, **128**, 1929-1957.

[17] Gärtner, J. (1988). On the McKean-Vlasov limit for interacting diffusions. *Math. Nachr.*, **137**, 197-248.

[18] Genon-Catalot, V. (1990). Maximum contrast estimation for diffusion processes from discrete observations, *Statistics*, **21**, 99-116.
[19] Genon-Catalot V. and Larédo, C. (2014). Asymptotic equivalence of nonparametric diffusion and Euler scheme experiments, *The Annals of Statistics*, **42**, 1145-1165.
[20] Genon-Catalot V. and Larédo, C. (2020). Probabilistic properties and parametric inference of small variance nonlinear self-stabilizing stochastic differential equation. Preprint Hal-02955171.
[21] Giesecke, K., Schwenkler, G. and Sirignano, J.A. (2020). Inference for large financial systems. *Mathematical Finance* **30**, 3-46.
[22] Gloter, A. and Sørensen, M. (2009). Estimation for stochastic differential equations with a small diffusion coefficient. *Stoch. Proc. Appl.* **119**, 679-699.
[23] Gobet, E., Hoffmann, M. and Reiss, M. (2004). Nonparametric estimation of scalar diffusions based on low frequency data. *The Annals of Statistics*, **32**, 2223-2253.
[24] Guy, R., Larédo, C. and Vergu, E. (2014). Parametric inference for discretely observed multidimensional diffusions with small diffusion coefficient. *Stoch. Proc. Appl.* **124**, 51-80.
[25] Herrmann, S., Imkeller, P. and Peithmann, D. (2008). Large deviations and a Kramers' type low for self-stabilizing diffusions. *The Annals of Applied Probability*, **18**, 1379-1423.
[26] Hirsch M.W. and Smale S. (1974). *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, New York San Francisco London.
[27] Hoffmann, M. (1999). Adaptive estimation in diffusion processes. *Stoch. Proc. and Appl.* **79**, 135-163.
[28] Höpfner, R., (2014). *Asymptotic Statistics with a view to stochastic processes.* Walter de Gruyter, Berlin/Boston.
[29] Iacus, S. M., (2010). *Simulation and inference for stochastic differential equations. With R examples.* Springer.
[30] Kessler, M., Lindner, A. and Sørensen, M., Editors (2012). *Statistical methods for stochastic differential equations.* CRC press. Taylor & Francis Group. Boca Raton.
[31] Kutoyants, Y.A., (1984). *Parameter estimation for stochastic processes.* Berlin: Heldermann.
[32] Kutoyants, Y.A., (2004). *Statistical inference for ergodic diffusion processes.* Springer, London.
[33] Larédo, C. (1990). A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process. *The Annals of Statistics* **18**, 1158-1171.
[34] Larédo, C. (2020). Statistical Inference for epidemic Processes in a Homogeneous Community. **Part IV** 363-472. Britton,T. and Pardoux, E., Editors. Lecture Notes in Mathematics 2255, Mathematical Biosciences Subseries. Springer.
[35] Kasonga, R.A. (1990). Maximum likelihood theory for large interacting systems. *SIAM Journal on Applied Mathematics*, **50**, 865-875.
[36] Mac Kean, H.P., Jr (1966). A class of Markov processes associated with nonlinear parabolic equation. *Proceedings NatL. Acad. Sci. U.S.A.* **56**, 1907-1911.
[37] Malrieu, F. (2003). Convergence to equilibrium for granular media equations and their Euler schemes. *Annals of applied Probability* **13**, 540-560.
[38] Masuda, H. (2007). Ergodicity and exponential beta-mixing for multidimensional diffusions with jumps. *Stoch. Proc. and Appl.*, **117**, 35-56.
[39] Masuda, H. (2019). Non-Gaussian quasi-likelihood estimation of SDE driven by locally stable Lévy process. *Stoch. Proc. Appl.* **129**, 1013-1059.
[40] Méléard, S. (1996). Asymptotic behaviour of some interacting particle systems; McKean-Vlasov and Boltzmann models. In *Probabilistic models for nonlinear partial differential equations*, Lecture Notes in Mathematics **1627**, 42-95, Springer.
[41] Molginer, A. and Edelstein-Keshet L. (1999). A non-local model for a swarm. *J. Math. Biol.* **38** 534-570.
[42] Piccini, U., de Gaetano, A. and Ditlevsen, S. (2010). Stochastic differential mixed-effects models. *Scand. J. Statist.* **37**, 67-90.
[43] Schmisser, E. (2014). Non-parametric adaptive estimation of the drift for a jump diffusion process. *Stoch. Proc. Appl.*, **124**, 883-914.
[44] Sørensen, M. and Uchida, M. (2003). Small diffusion asymptotics for discretely sampled stochastic differential equations. *Bernoulli* **9**, 1051-1069.
[45] Sznitman, A.-S. (1991). Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX-1989. Lecture Notes in Math.* **1464**, 165-251. Springer, Berlin.
[46] Yoshida, N. (1992). Estimation for diffusion processes from discrete observation. *J. Multivariate Analysis*, **41**, 220-242.

[47] Yoshida, N. (1992). Asymptotic expansions of maximum likelihood estimators for small diffusions via the theory of Malliavin-Watanabe. *Probability Theory and Related Fields*, **92**, 275-311.